
GENERACIÓN DE RECURSOS LINGÜÍSTICOS MEDIANTE LA EXTRACCIÓN DE RELACIONES ENTRE CONCEPTOS

Paloma Galván Calleja

Máster en Ingeniería Informática. Facultad de Informática
Universidad Complutense de Madrid



Trabajo de Fin de Máster en Ingeniería Informática
2016-2017

Directores:

Gonzalo Méndez Pozo
Raquel Hervás Ballesteros
Virginia Francisco Gilmartín

GENERACIÓN DE RECURSOS LINGÜÍSTICOS MEDIANTE LA EXTRACCIÓN DE RELACIONES ENTRE CONCEPTOS

Trabajo de Fin de Máster en Ingeniería Informática
Departamento de Ingeniería del Software e Inteligencia Artificial

Autor
Paloma Galván Calleja

Directores
Gonzalo Méndez Pozo
Raquel Hervás Ballesteros
Virginia Francisco Gilmartín

Convocatoria: *Septiembre 2017*
Calificación: *8*

Máster en Ingeniería Informática
Facultad de Informática
Universidad Complutense de Madrid

21 de septiembre de 2017

Autorización de difusión

La abajo firmante, matriculada en el Máster en Ingeniería Informática de la Facultad de Informática, autoriza a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor el presente Trabajo de Fin de Máster: GENERACIÓN DE RECURSOS LINGÜÍSTICOS MEDIANTE LA EXTRACCIÓN DE RELACIONES ENTRE CONCEPTOS, realizado durante el curso académico 2016-2017 bajo la dirección de Gonzalo Méndez Pozo, Raquel Hervás Ballesteros y Virginia Francisco Gilmartín en el Departamento de Ingeniería del Software e Inteligencia Artificial, y a la Biblioteca de la UCM a depositarlo en el Archivo Institucional E-Prints Complutense con el objeto de incrementar la difusión, uso e impacto del trabajo en Internet y garantizar su preservación y acceso a largo plazo.

Paloma Galván Calleja

21 de septiembre de 2017

Agradecimientos

Quiero agradecer a mis directores Gonzalo, Raquel y Virginia por la paciencia que han tenido para guiarme en este trabajo y también, por la dedicación y esfuerzo que emplean en su labor docente y de investigación, es admirable la forma en que consiguen compaginarlo todo. También agradezco al grupo NIL al completo haberme permitido vivir esta experiencia tan enriquecedora y haberme mostrado toda su ayuda y apoyo durante este tiempo.

Este trabajo se ha realizado en el marco de los proyectos ConCreTe e IDiLyCo. El soporte financiero del proyecto ConCreTe parte del programa Future and Emerging Technologies (FET) que se engloba en el Seventh Framework Programme for Research of the European Commission, con la subvención FET número 611733. El proyecto IDiLyCo está subvencionado por el Ministerio de Economía y Competitividad Español con la subvención número TIN2015-66655-R.

Resumen

Hoy en día disponemos de una gran cantidad de información en Internet y sería muy positivo poder obtener beneficio de ella para desarrollar sistemas de generación de texto. Nuestro objetivo inicial es desarrollar un generador de recursos lingüísticos basado en asociaciones de conceptos. Un generador de lenguaje natural normalmente tiene acceso a un gran conjunto de conocimiento del que selecciona información y presenta en diferentes formas a los usuarios finales. Por lo tanto, un sistema generador de lenguaje natural debe escoger información de la base de conocimiento, decidir cómo organizarla, y determinar cómo producir el texto en lenguaje natural.

El problema ante el que nos encontramos en muchas ocasiones es la desorganización de tanta información disponible en Internet, pues para que resulte útil en el desarrollo de sistemas de Generación de Lenguaje Natural, la información debe estar organizada y relacionada. Esto se está cambiando con la llegada de herramientas relacionadas con la Web Semántica. En este trabajo se va a implementar un generador que, a partir de la información obtenida de algunas de estas herramientas, aquellas relacionadas con la asociación de conceptos semánticos, desarrolle recursos lingüísticos que serán posteriormente evaluados. El objetivo final de este trabajo es conocer el valor de las asociaciones de palabras para la generación de diferentes recursos de lenguaje natural.

En un primer momento se ha desarrollado un generador de adivinanzas que crea acertijos sobre conceptos representados como sustantivos comunes. Estos acertijos se basan en comparaciones entre el concepto destino (por ejemplo, un diamante) y otras entidades que comparten algunas de sus propiedades (por ejemplo, el cemento). Los acertijos resultantes se componen como una secuencia de comparaciones utilizando esta plantilla: “Qué es... tan *atributo* como *concepto*?”, donde “atributo” es una propiedad del concepto buscado que es la respuesta al acertijo, y “concepto” es una entidad diferente que comparte el valor del atributo con el concepto destino. Por ejemplo, “¿Qué es... tan *duro* como el *cemento* y tan *transparente* como el *pelo*?” es un acertijo generado para el concepto diamante por el generador de adivinanzas.

Posteriormente, se analiza la relación entre asociaciones de palabras ya disponibles en diferentes recursos web y el tipo de figuras retóricas que las personas emplean regularmente. El objetivo es identificar regularidades o heurísticas que pueden ser empleadas en una etapa posterior para desarrollar un generador de texto. Será estudiada la capacidad de este sistema para la generación automática de analogías, símiles y metáforas.

Palabras clave

Creatividad, Recursos lingüísticos, Asociación, Acertijos, Figuras retóricas, Metáfora, Símil, Analogía

Abstract

Nowadays we have a lot of information on the Internet and it would be very positive to be able to benefit from it to develop systems of text generation. Our initial goal is to develop a language resource generator based on concept associations. A natural language generator usually has access to a large set of knowledge from which it selects information and presents it to end users in different forms. Therefore, a natural language generator system must choose information from the knowledge base, decide how to organize it, and determine how to produce text in natural language.

The problem we are faced with on many occasions is the disorganization of so much information available on the Internet, because to be useful in the development of Natural Language Generation systems, the information must be organized and related. This is changing with the arrival of tools related to the Semantic Web. In this work we will implement a generator that, based on the information obtained from some of these tools, those related to the association of semantic concepts, develop linguistic resources that will be further evaluated. The final aim of this work is to know the value of word associations for the generation of different natural language resources.

Initially a riddle generator has been developed that creates riddles about concepts represented as common nouns. These riddles are based on comparisons between the target concept (i.e. a diamond) and other entities that share some of their properties (i.e. concrete). The resulting riddles are composed as a sequence of comparisons using this template: “What is ... as *attribute* as *concept*?”, where “attribute” is a property of the searched concept that is the answer to the riddle, and “concept” is a different entity that shares the value of the attribute with the target concept. For example, “What is ... as *hard* as *concrete* and as *transparent* as *hair*?” is a riddle generated for the concept diamond by the riddle generator.

Later, the relation between words associations already available in different web resources and the type of rhetorical figures that people use regularly is analysed. The aim is to identify regularities or heuristics that can be used in a posterior stage to develop a text generator. The capacity of this system will be studied for the

automatic generation of analogies, similes and metaphors.

Keywords

Creativity, Linguistic Resources, Word association, Riddles, Tropes, Rhetorical figures, Metaphor, Simile, Analogy

Índice

| | |
|--|-----------|
| 1. Introducción | 1 |
| 1.1. Creatividad computacional | 2 |
| 1.1.1. ConCreTe | 4 |
| 1.1.2. Mi papel | 5 |
| 1.2. Motivación | 6 |
| 1.3. Estructura del documento | 6 |
| 2. Introduction | 9 |
| 2.1. Computational creativity | 10 |
| 2.1.1. ConCreTe | 12 |
| 2.1.2. My role | 12 |
| 2.2. Motivation | 13 |
| 2.3. Document structure | 14 |
| 3. Objetivos | 15 |
| 3.1. Científicos | 17 |
| 3.2. Tecnológicos | 17 |
| 3.3. Método de trabajo | 18 |
| 4. Estado de la cuestión | 21 |
| 4.1. Conceptos lingüísticos: Figuras retóricas | 21 |
| 4.2. Herramientas de asociación de conceptos | 22 |
| 4.2.1. Thesaurus Rex | 23 |
| 4.2.2. ConceptNet | 25 |
| 4.2.3. WordNet | 25 |
| 4.3. Generación de acertijos | 26 |
| 4.3.1. Adivinanzas: accesibilidad y representación de la información | 26 |
| 4.3.2. JAPE | 26 |
| 4.3.3. Generación automática de acertijos | 27 |

| | | |
|-----------|---|-----------|
| 4.3.4. | Utilizando modelos temáticos y medidas de similitud semántica | 28 |
| 4.3.5. | TheRiddlerBot | 28 |
| 4.4. | Generación de figuras retóricas | 29 |
| 4.4.1. | Aproximaciones a las figuras retóricas en GLN | 30 |
| 4.4.2. | Aproximaciones para la construcción conceptual de figuras retóricas | 30 |
| 4.4.3. | Jigsaw Bard | 31 |
| 4.4.4. | Thesaurus Rex | 31 |
| 4.4.5. | Metaphor Magnet | 33 |
| 4.4.6. | Metaphor Eyes | 33 |
| 4.4.7. | FIGURE8 | 34 |
| 5. | Proceso de extracción de relaciones entre conceptos | 35 |
| 5.1. | Proceso de asociación de palabras | 35 |
| 5.1.1. | Entrada | 35 |
| 5.1.2. | Proceso | 36 |
| 5.1.3. | Salida | 39 |
| 5.2. | Servicio web | 39 |
| 6. | Generación de adivinanzas a partir de relaciones | 43 |
| 6.1. | Generación de adivinanzas | 43 |
| 6.2. | Evaluación del generador de adivinanzas | 45 |
| 6.2.1. | Diseño de la evaluación | 46 |
| 6.2.2. | Resultados de la evaluación | 47 |
| 6.2.3. | Discusión | 48 |
| 6.3. | Conclusiones | 50 |
| 7. | Generación de figuras retóricas a partir de relaciones | 53 |
| 7.1. | Generación de figuras retóricas | 54 |
| 7.2. | Evaluación del generador de figuras retóricas | 55 |
| 7.2.1. | Diseño de la evaluación | 56 |
| 7.2.2. | Resultados de la evaluación | 57 |
| 7.2.3. | Discusión | 61 |
| 7.3. | Conclusiones | 62 |
| 8. | Conclusiones y trabajo futuro | 65 |
| 8.1. | Conclusiones | 65 |
| 8.2. | Trabajo futuro | 67 |

| | |
|---|-----------|
| 9. Conclusions and future work | 71 |
| 9.1. Conclusions | 71 |
| 9.2. Future work | 73 |
| Bibliografía | 75 |
| A. Evaluación de acertijos | 81 |
| B. Evaluación de figuras retóricas | 85 |
| C. Artículo publicado para LREC 2016 | 89 |
| D. Artículo publicado para ICC 2016 | 97 |

Índice de figuras

| | |
|--|----|
| 4.1. Captura del servicio web Thesaurus Rex para el concepto <i>coffee</i> . . | 24 |
| 4.2. Captura de la página de TheRiddlerBot en Twitter | 29 |
| 4.3. Captura de la aplicación Thesaurus Rex, para la búsqueda de conceptos <i>divorce and war</i> | 32 |
| 5.1. Diagrama de flujo del proceso | 38 |
| 6.1. Captura del servicio web Thesaurus Rex de las categorías a las que pertenece el atributo <i>stellar</i> | 50 |
| 6.2. Captura del servicio web Thesaurus Rex de las categorías a las que pertenece el atributo <i>fast</i> | 50 |
| 7.1. Gráfico de barras con los resultados de los conceptos abstractos de las figuras retóricas | 60 |
| 7.2. Gráfico de barras con los resultados de los conceptos concretos de las figuras retóricas | 61 |
| 7.3. Gráfico de barras con los resultados totales de las figuras retóricas | 61 |
| 8.1. Resultados positivos y negativos del concepto <i>sun</i> en Metaphor Magnet | 69 |

Índice de tablas

| | |
|--|----|
| 4.1. Ejemplos de categorías (con sus pesos) para el concepto <i>coffee</i> en Thesaurus Rex | 23 |
| 4.2. Ejemplos de propiedades (con sus pesos) para el concepto <i>coffee</i> en Thesaurus Rex | 23 |
| 5.1. Ejemplos de asociaciones de palabras obtenidas. Las palabras en negrita representan las elecciones hechas para cada ejemplo. | 36 |
| 6.1. Ejemplos de comparaciones obtenidas para ser parte de las adivinanzas. Las palabras en negrita representan las elecciones hechas para cada ejemplo. | 44 |
| 6.2. Adivinanza A6 de la evaluación a partir del concepto <i>diamond</i> . . . | 47 |
| 6.3. Porcentaje de acierto para cada adivinanza en el conjunto aleatorio | 47 |
| 6.4. Porcentaje de acierto para cada adivinanza en el conjunto revisado | 48 |
| 6.5. Porcentaje de adivinanzas acertadas | 48 |
| 7.1. Ejemplos de asociaciones obtenidas para formar las figuras retóricas. Las palabras en negrita representan las elecciones hechas para cada ejemplo. | 54 |
| 7.2. Resultados de metáforas | 58 |
| 7.3. Resultados de símiles | 58 |
| 7.4. Resultados de analogías | 58 |
| 7.5. Resultados generales de la evaluación | 58 |
| A.1. Acertijos aleatorios utilizados en la evaluación | 82 |
| A.2. Acertijos revisados utilizados en la evaluación | 83 |

Capítulo 1

Introducción

Hoy en día disponemos de una cantidad ingente de información en Internet, facilitada por enciclopedias en línea, periódicos, redes sociales, blogs, foros de discusión, o cualquier página web ya sea de investigación, divulgativa, empresarial, personal, etc. Un aspecto positivo sería obtener beneficio de dicha cantidad de información disponible en plataformas en línea para desarrollar sistemas de generación de texto. El problema ante el que nos encontramos es la desorganización de tanta información, pues para que resulte útil en el desarrollo de sistemas de Generación de Lenguaje Natural, la información debe estar organizada y relacionada. Por este motivo, para el desarrollo de muchas herramientas de generación de texto, se utilizan bases de datos con la información representada en forma de redes de conceptos con sus relaciones, como son mapas conceptuales u ontologías.

Como indica Jordán (1992), la Generación de Lenguaje Natural es el proceso de la construcción de expresiones, o partes de ellas, tanto habladas como escritas en lenguaje natural para la comunicación con fines específicos. El principal énfasis no es sólo el facilitar el uso del ordenador sino también construir herramientas para extender, aclarar y verificar teorías que se han formulado en lingüística, psicología y sociología acerca de la comunicación entre humanos. Un generador de lenguaje natural normalmente tiene acceso a un gran conjunto de conocimiento del que selecciona información y presenta en diferentes formas a los usuarios finales. Por tanto, la generación de texto es un problema de toma de decisiones con las restricciones, entre otras, de un conocimiento proposicional, herramientas lingüísticas disponibles y los objetivos de comunicación del usuario a quien se dirige el texto. Un sistema generador de texto debe escoger información de la base de conocimiento, decidir cómo organizarla, y determinar cómo producir el texto en lenguaje natural. Esto incluye el decidir acerca de la entrada del léxico y de las estructuras sintácticas, con lo cual, el proceso de generación se divide en una componente de planificación del texto y otra de generación propiamente dicha.

Hace unos años, las técnicas para la representación del conocimiento no estaban suficientemente desarrolladas para abarcar niveles aceptables de eficacia dadas grandes cantidades de conocimiento, como sería necesario para entender una gran variedad de temas. Existían sistemas de demostración que operaban con una cantidad de conocimiento limitada, mas bien de artesanía, diseñados para procesar un conjunto pequeño de entradas en forma de mapas conceptuales. Con la llegada de la Web Semántica se ha hecho menos necesario el uso de estas herramientas para el modelado de dominios. La idea de la Web Semántica se basa en añadir metadatos semánticos y ontológicos a la World Wide Web de manera formal, de modo que sea posible procesarlos automáticamente por las máquinas, permitiendo compartir y utilizar la información de la comunidad.

En la actualidad, tecnologías como la Inteligencia Artificial están cambiando la forma en que concebimos la creatividad. Los programas están ejerciendo importantes labores en tareas creativas como la arquitectura, el dibujo, la música o la poesía, en las artes y ciencias en general (véase la sección 1.1). Se está empezando a pensar en los computadores no sólo como una herramienta de ayuda para crear, sino como una entidad creativa en sí misma, considerándolos como plumas, pinceles, lienzos o instrumentos musicales. Esta perspectiva ha originado la creatividad computacional, un nuevo subcampo de la Inteligencia Artificial. Este trabajo aborda la posibilidad de conseguir creatividad computacional mediante un sistema de asociación de palabras en el campo de la Generación de Lenguaje Natural.

1.1. Creatividad computacional

Uno de los aspectos más estudiados de la conducta humana es la creatividad, la cual nos permite reinventar lo antiguo e idear lo nuevo. Profesionales de diversas áreas se han unido bajo el emergente campo de investigación de la creatividad computacional. Este campo multidisciplinario aúna elementos de inteligencia artificial, ciencia cognitiva, psicología y filosofía. La creatividad computacional estudia el potencial de las computadoras de ser autónomamente creativas, pero también nos permite entender la creatividad y producir programas que sirvan como colaboradores creativos de las personas. A través del entendimiento del comportamiento creativo de los seres humanos, se formulan algoritmos y se diseñan programas que utilizando una computadora puedan simular la creatividad humana o reproducirla en forma de historias, cuentos, figuras retóricas, poemas, bromas, adivinanzas, teoremas, canciones, pinturas, juegos, etc. El trabajo teórico sobre la naturaleza y adecuada definición de la creatividad se realiza en paralelo con el trabajo práctico de implementar sistemas que muestran creatividad, de modo que informándose de los avances ambas líneas de trabajo se enriquecen mutuamente.

En el ámbito de la música, la creatividad computacional ha desarrollado tanto partituras musicales para la actuación de músicos humanos como para la reproducción por computador. David Cope desarrolló un sistema llamado *Experiments in Musical Intelligence* (Cope, 1987), capaz de analizar música creada por un compositor humano para generar composiciones musicales nuevas del mismo estilo. Iamus (Díaz-Jerez, 2011) es el primer equipo que compone música a partir de cero, concretamente música clásica contemporánea, y genera obras que pueden tocar intérpretes profesionales, como sucedió con la orquesta sinfónica de Londres. Otro gran logro a nivel musical lo ha conseguido el robot Shimon (Weinberg et al., 2009), una herramienta interactiva que ha logrado demostrar improvisación componiendo jazz.

En la generación de arte visual ha habido éxitos tanto en arte abstracto como en arte representativo. El programa más notable es AARON de Harold Cohen (Cohen, 1995), que es capaz de generar dibujos en blanco y negro o en color y cuyas obras tienen la calidad suficiente para exhibirse en galerías de renombre. El sistema NEvAr de Penousal Machado (Machado y Cardoso, 2000) genera una superficie tridimensional de color utilizando un algoritmo genético. Por su parte, Simon Colton (Colton, 2012) desarrolló el sistema *The Painting Fool* que dibuja imágenes digitales de una escena dada en diferentes estilos de pintura, paletas de colores y tipos de pincel.

El idioma es un área que ofrece enormes oportunidades para la creatividad, como juegos de palabras, rimas, sarcasmo, ironía, metáforas, símiles, analogías, chistes, etc. Hablantes nativos crean habitualmente frases fáciles de entender que dan nuevo significado a diferentes conjuntos de palabras, aunque nunca aparecerán en un diccionario. El campo de la Generación de Lenguaje Natural ha sido ampliamente estudiado:

- *Chistes*. Los sistemas de generación de bromas con más éxito se han centrado en la generación de juegos de palabras. El sistema JAPE, parte del trabajo de Binsted y Ritchie (1997); Ritchie (2003), demuestra que puede generar una amplia variedad de juegos de palabras, los cuales fueron evaluados por niños como novedosos y cómicos. El sistema STANDUP (Waller et al., 2009) se ha desarrollado como una versión mejorada de JAPE, orientado como un medio para mejorar la interacción lingüística de niños con discapacidades de comunicación. En la generación de bromas que implican otros aspectos del lenguaje natural, destaca la obra de Hans Wim Tinholt y Anton Nijholt (Tinholt y Nijholt, 2007), que utiliza un malentendido deliberado causado por la ambigüedad en las referencias pronominales, así como el sistema HAHAAcronym de Stock y Strapparava (2003) que desarrolla la generación de acrónimos humorísticos.

- *Poesía*. Al igual que sucede con los chistes, la poesía requiere la interacción de diferentes y complejos requisitos. No hay generadores de poesía de propósito general que logren aunar dichos requisitos adecuadamente, como son significado, estructura y formato de la rima, pero el sistema ASPERA desarrollado por Gervás (2001) compone poesía formal en español de forma interactiva semiautomática. ASPERA combina en el proceso de composición la Generación de Lenguaje Natural y las técnicas de razonamiento basado en casos (CBR) para generar formulaciones poéticas de un texto de entrada dado.
- *Historias*. James Meehan, ya en los 70, desarrolló TALE-SPIN (Meehan, 1977) que creaba historias en forma de descripciones narrativas para la resolución de un problema, fijando un objetivo para los personajes de la historia de modo que su búsqueda de la solución quedase registrada. Por otra parte, el sistema MINSTREL (Turner, 1993) representa una elaboración más compleja de este enfoque, diferenciando objetivos a nivel de personajes en la historia y objetivos a nivel de autor para la historia. El modelo MEXICA (Pérez Y Pérez y Sharples, 2001) se interesa más por el proceso creativo de la narración, e implementa un modelo cognitivo de la escritura creativa.
- *Metáforas, símiles y analogías*. La interpretación o procesamiento de las metáforas y los símiles a nivel computacional se ha estudiado como un proceso basado en el conocimiento. Tony Veale y Yanfen Hao han desarrollado el sistema Sardonicus (Veale y Hao, 2007), que obtiene su base de datos de símiles a partir de la web, los cuales se clasifican como objetivos o irónicos. Estos símiles se utilizan como la base del sistema de generación de metáforas Aristotle (Veale, 2011), el cual puede sugerir metáforas léxicas para un objetivo descriptivo dado.

1.1.1. ConCreTe

Este trabajo fin de máster se ha realizado en el marco del proyecto europeo ConCreTe (*Concept Creation Technology*, FET grant number 611733¹) bajo el grupo de investigación NIL del Departamento de Ingeniería del Software e Inteligencia Artificial. Este proyecto se ha realizado en colaboración con universidades en Londres, Coímbra, Helsinki, Liubliana, Madrid y Enschede. El proyecto tiene como objetivo estudiar la creatividad conceptual en seres humanos y en máquinas. También ofrece una visión de sistemas que pueden comportarse de manera comparable a la creatividad humana, con una mejor interacción entre humano y máquina, mejores sistemas autónomos y nuevos tipos de creatividad.

¹<http://conceptcreationtechnology.eu/>

El software del proyecto se encuentra en línea a través de una plataforma de desarrollo llamada ConCreTe Flows². Con esta plataforma se trabaja en entornos de aprendizaje creativos en línea, para enseñar o apoyar actividades creativas y promover la creatividad. Se utiliza la tecnología de la Web Semántica para evitar el modelado de dominios. En esta plataforma de desarrollo e intercambio se pueden activar diferentes herramientas para construir flujos de trabajo, como por ejemplo dado el contenido en forma de texto de dos fuentes de información en línea puede producir un mapa conceptual de cada uno y además mezclar la información de manera gráfica, textual o visual. Otras tareas muy interesantes son la generación de poesía, la generación de letras de canciones o la generación de resúmenes.

1.1.2. Mi papel

Este proyecto Europeo abarca gran parte de las disciplinas dentro de la creatividad computacional, pero mi lugar se ha centrado en la Generación de Lenguaje Natural, más concretamente adivinanzas y figuras retóricas.

La cantidad de información disponible hoy en día en Internet ofrece un enorme abanico de posibilidades. Debido a que ya existen aplicaciones que relacionan esta información, se hace más sencillo crear aplicaciones especializadas en generación de lenguaje. Dado que se deseaba estudiar la repercusión de las asociaciones de conceptos en la generación de recursos lingüísticos, se buscaron tipos de recursos en los que las asociaciones de conceptos de alguna manera jugaran un papel importante en su creación. Una de las opciones que parecía interesante era desarrollar una aplicación que presentase adivinanzas en forma de analogía, y de este modo, estudiar la relación entre los conceptos a través de propiedades compartidas para ayudarnos a generar mejores acertijos.

Posteriormente, en base al razonamiento de seguir estudiando las asociaciones de palabras, parecía motivadora la idea de crear figuras retóricas a partir de las ideas o conceptos relacionados también utilizados para la generación de adivinanzas. Las figuras retóricas que se eligieron fueron metáfora, símil y analogía por ser las que a través de un desplazamiento de propiedades crean asociaciones entre conceptos para embellecer un texto. Esta parte resultaba más experimental, queríamos comprobar cómo de útil podía ser nuestro proceso de asociación de conceptos para generar lenguaje figurativo y compararlo con el generado por personas.

²<http://concreteflows.ijs.si/>

1.2. Motivación

Cuando hablamos de bases de conocimiento donde los conceptos tienen propiedades asociadas, se nos presenta una gran cantidad de información comparativa. Esta información se encuentra implícita en las propiedades que comparten estos conceptos y esta relación crea una asociación entre ambos conceptos. Por lo tanto, toda esta información puede ser útil en las tareas donde es necesario establecer relaciones entre conceptos de forma automática, tales como la generación de comparaciones entre entidades basadas en las propiedades compartidas (por ejemplo, *esta camisa es blanca como la nieve*). El trabajo descrito es este documento pretende encontrar, a través de búsquedas con herramientas ya disponibles, asociaciones de conceptos y comprobar qué valor pueden tener en la Generación de Lenguaje Natural. Estas tareas se incluyen en el campo de la creatividad computacional, tratando de simular el comportamiento natural de los seres humanos de ser creativos utilizando programas informáticos.

El principal motivo de la elección del tema de este trabajo, por tanto, fue la vinculación con la creatividad computacional y con el proyecto ConCreTe. Para empezar mi aprendizaje y entendimiento de las materias relacionadas con este proyecto tuve que leer y estudiar distintos artículos. En un primer momento, comencé a informarme sobre algunas herramientas que luego me han resultado muy útiles y en base a la utilización de estas herramientas vinculadas con la Web Semántica, fue tomando forma el sistema que ha sido finalmente desarrollado.

En un principio se ha desarrollado un generador de adivinanzas que crea acertijos sobre conceptos representados como sustantivos comunes. Estos acertijos se basan en comparaciones entre un concepto destino (i.e. *esta camisa*) y otras entidades que comparten algunas de sus propiedades (i.e. *la nieve*).

Posteriormente, se explora la relación entre asociaciones de palabras encontradas a través de recursos semánticos ya disponibles y el tipo de figuras retóricas que las personas emplean regularmente. El objetivo es identificar posibles regularidades, intuiciones o heurísticas que pueden ser empleadas en una etapa posterior para desarrollar un generador de texto capaz de utilizar este tipo de figura retórica. El potencial de este sistema será estudiado para la generación automática de analogías, símiles y metáforas.

1.3. Estructura del documento

Este documento sigue una estructura similar a la que se ha seguido durante el desarrollo del proyecto.

En primer lugar, en el Capítulo 3 se proponen los objetivos de este proyecto y

cuáles son las pautas para conseguirlos.

En el Capítulo 4 se hace un repaso de las tecnologías usadas para el desarrollo del sistema, así como de proyectos relacionados. Proporciona una visión general del estado de arte tanto en generación de acertijos como en generación de lenguaje figurativo.

Una vez conocido el contexto de la generación de lenguaje y las tecnologías, en el Capítulo 5 se explica el primer objetivo en el desarrollo de este proyecto, introduciendo el proceso de asociación de palabras llevado a cabo con la herramienta Thesaurus Rex.

En el Capítulo 6 se muestra nuestro enfoque para generar acertijos utilizando asociaciones de palabras. Se presenta una primera evaluación y los resultados obtenidos a partir de una serie de cuestionarios donde las personas trataron de resolver dos conjuntos de adivinanzas generados por nuestro sistema. También contiene algunas conclusiones y el trabajo futuro sobre la generación de acertijos, así como la implementación de alguna de ellas.

A continuación, el Capítulo 7 muestra una segunda evaluación para el proceso de asociación de palabras, pero esta vez utilizado para la generación de figuras retóricas. Se detallan tanto el diseño como los resultados, las conclusiones obtenidas y el trabajo futuro que se podría llevar a cabo sobre esta aplicación.

Por último, en los Capítulos 8 y 9 se presentan tanto las conclusiones a las que se han llegado como las características o modificaciones que se podrían añadir en futuras iteraciones del proyecto.

Chapter 2

Introduction

Nowadays we have a huge amount of information on the Internet, provided by online encyclopaedias, newspapers, social networks, blogs, discussion forums, or any web page whether research, information, business, personal, etc. A positive aspect would be to obtain benefit of such amount of information available on online platforms to develop text generation systems. But we must face the problem of the disorganization of so much information, because the information must be organized and related to be useful. For this reason, the development of many tools of text generation requires the use of databases with information represented in the form of networks of concepts with their relations, such as conceptual maps or ontologies.

As indicated by Jordán (1992), Natural Language Generation is the process of constructing expressions, or parts of them, both spoken and written in natural language for communication with specific purposes. The main emphasis is not only to facilitate the use of the computer but also to build tools to extend, clarify and verify theories that have been formulated in linguistics, psychology and sociology about human communication. A natural language generator usually has access to a large set of knowledge from which it selects information and presents it to the end users in different forms. Therefore, text generation is a problem of decision making with the restrictions, among others, of a propositional knowledge, available linguistic tools and the communication objectives of the user to whom the text is directed. A text-generating system must choose information from the knowledge base, decide how to organize it, and determine how to produce the text in natural language. This includes deciding on the input of the lexicon and syntactic structures, whereby the generation process is divided into a text planning component and a generation component itself.

A few years ago, techniques for knowledge representation were not sufficiently developed to encompass acceptable levels of efficacy given large amounts of knowledge, as would be required to understand a wide variety of topics. There were

demonstration systems that operated with a limited amount of knowledge, rather craftsmanship, designed to process a small set of inputs in the form of concept maps. The use of these tools for domain modelling has become less necessary with the advent of the Semantic Web. The idea of the Semantic Web is based on adding semantic and ontological metadata to the World Wide Web in a formal way, so machines are capable to process them automatically, allowing to share and use the information of the community.

Today, technologies such as Artificial Intelligence are changing the way we conceive creativity dramatically. Programs are performing important work in creative tasks such as architecture, drawing, music or poetry, in the arts and sciences in general (see section 1.1). It is beginning the thought of computer not only as an aid tool to create, but as a creative entity itself, considering them as pens, brushes, canvas or musical instruments. This perspective has originated a new subfield of Artificial Intelligence called computational creativity. This work addresses the possibility of achieving computational creativity through a word association system in the field of Natural Language Generation.

2.1. Computational creativity

One of the most studied aspects of human behaviour is creativity, which allows us to reinvent what it is old and devise what it is new. Professionals from diverse areas have come together under the emerging research field of computational creativity. This multidisciplinary field combines elements of artificial intelligence, cognitive science, psychology and philosophy. Computational creativity studies the potential of computers to be autonomously creative, but it also allows us to understand creativity and produce programs that serve as creative collaborators of people. Though the understanding of the creative behaviour of human beings, algorithms are formulated and programs are designed that can simulate human creativity using a computer or can reproduce it in the form of stories, tales, rhetorical figures, poems, jokes, riddles, theorems, songs, paintings, games, etc. The theoretical work on the adequate definition of creativity is performed in parallel with the practical work of implementing systems that show creativity, and in the way, they show the advances, both lines of work can be mutually enriched.

In the field of music, computational creativity has developed musical scores for the performance of human musicians and for computer reproduction. David Cope developed a system called Experiments in Musical Intelligence (Cope, 1987), able to analyse music created by a human composer to generate new musical compositions of the same style. Iamus (Diaz-Jerez, 2011) is the first team to compose music from scratch, specifically contemporary classical music, and generates works

that can play professional performers, as was the case with the London Symphony Orchestra. Another great achievement at the musical level has been achieved by the robot Shimon (Weinberg et al., 2009), an interactive tool that has managed to demonstrate improvisation composing jazz.

In the generation of visual art there have been successes both in abstract art and in representative art. The most notable program is AARON of Harold Cohen (Cohen, 1995), which can generate black and white or colour drawings and whose works have enough quality to be displayed in renowned galleries. NEvAr system of Penousal Machado (Machado y Cardoso, 2000) generates a three-dimensional surface of colour using a genetic algorithm. Meanwhile, Simon Colton (Colton, 2012) developed The Painting Fool system that draws digital images of a scene given in different painting styles, colour palettes and brush types.

Language is an area that offers enormous opportunities for creativity, such as puns, rhymes, sarcasm, irony, metaphors, similes, analogies, jokes, etc. Native speakers usually create easy-to-understand sentences that give new meaning to different sets of words, though they will never appear in a dictionary. The field of Natural Language Generation has been extensively studied:

- *Jokes.* The most successful joke generation systems have focused on the generation of puns. JAPE system, part of the work of Binsted y Ritchie (1997); Ritchie (2003), demonstrates that can generate a wide variety of puns, which were evaluated by children as novel or comic. STANDUP system (Waller et al., 2009) has been developed as an improved version of JAPE, oriented to improve the linguistic interaction of children with communication disabilities. In joke generation, which involves other aspects of natural language, the work of Hans Wim Tinholt and Anton Nijholt (Tinholt y Nijholt, 2007) highlights using a deliberate misunderstanding caused by the ambiguity in pronominal references. HAHAAcronym system of Stock y Strapparava (2003) that develops the generation of humorous acronyms is also underlined.
- *Poetry.* As with jokes, poetry requires the interaction of different and complex requirements. General purpose poetry generators fail to combine these requirements adequately, such as meaning, structure and format of rhyme, but ASPERA system developed by Gervás (2001) composes Spanish poetry in an interactive semi-automatic way. ASPERA combines in the composition process, Natural Language Generation and the techniques of case-based reasoning (CBR) to generate poetic formulations of a given input text.
- *Stories.* James Meehan developed TALE-SPIN (Meehan, 1977), that creates stories as narrative descriptions for solving a problem, fixing a target for the story characters so that their search for the solution is recorded. On the other

hand, MINSTREL system (Turner, 1993) represents a more complex elaboration of this approach, differentiating character-level objectives of the story and author-level goals. MEXICA model (Pérez Y Pérez y Sharples, 2001) is more concerned in the creative process of storytelling, and implements a cognitive model of creative writing.

- *Metaphors, similes and analogies.* The interpretation or processing of metaphors and similes at the computational level has been studied as a process based on knowledge. Tony Veale and Yanfen Hao have developed Sardonicus system (Veale y Hao, 2007), which obtains its database of similes from the web, which are classified as objective or ironic. These similes are used as the basis of the metaphor generation system Aristotle (Veale, 2011), which may suggest lexical metaphors for a given descriptive purpose.

2.1.1. ConCreTe

This master's project has been conducted under the European project ConCreTe (Concept Creation Technology, FET grant number 611733¹) under NIL research group of Departamento de Ingeniería del Software e Inteligencia Artificial.

This project has been carried out in collaboration with universities in London, Coimbra, Helsinki, Ljubljana, Madrid and Enschede. The project aims to study conceptual creativity in humans and machines. It also offers a vision of systems that can behave in ways comparable to human creativity, with a better interaction between human and machine, better autonomous systems and new types of creativity.

The project software is online through a development platform called ConCreTe Flows². With this platform, it is possible to work in creative learning environments online, to teach or support creative activities and promote creativity. Semantic Web technology is used to avoid domain modelling. In this development and exchange platform, different tools can be activated to build workflows, such as given the content of two information sources online in text form, a conceptual map of each can be produced and the information can be mixed in a graphic, textual or visual way. Other very interesting tasks are poetry, lyrics or summary generation.

2.1.2. My role

This European project covers a large part of the disciplines within computational creativity, but my role has been focused on Natural Language Generation, more specifically riddles and rhetorical figures.

¹<http://conceptcreationtechnology.eu/>

²<http://concreteflows.ijs.si/>

The amount of information available today on the Internet offers a huge range of possibilities. Because there are already applications that relate this information, it makes easier to create specialized applications in language generation. Since we wanted to study the value of concept associations in the generation of linguistic resources, we searched types of resources in which concept associations somehow played an important role in their creation. One of the options that seemed interesting was to develop an application that would present riddles in the form of analogies, and thus, to study the relationship between concepts through shared properties to help us generate better riddles.

Subsequently, based on the reasoning of further study of word associations, the idea of creating rhetorical figures from the related ideas or concepts also used to generate riddles seemed motivating. The rhetorical figures that were chosen were metaphor, simile and analogy for being those that through a displacement of properties create associations between concepts to beautify a text. This part was more experimental, we wanted to check how useful could be the association of concepts to generate figurative language by a system and compare it with the one generated by people.

2.2. Motivation

In knowledge bases where concepts have associated properties, there is a large amount of comparative information. This information is implicitly encoded in the values of the properties that these concepts share and this relation creates an association between both concepts. Therefore, all this information can be useful in tasks where it is required to automatically establish relations between concepts, such as the generation of comparisons between entities based on shared properties (e.g. *this shirt is as white as snow*). The work described in this paper aims to find, through searches with tools already available, associations of concepts and to check what value they can have in the Natural Language Generation. These tasks are included in computational creativity, trying to simulate the natural behaviour of human beings to be creative using computer programs.

Therefore, the main reason for the choice of this work subject was the link with the computational creativity and with project ConCreTe. To begin my learning and understanding of the subjects related to this project, I had to read and study different papers. At first, I began to learn about some tools that later I have found very useful. With the use of these tools linked with Semantic Web, the system that has been finally developed was taking shape.

Initially, a riddle generator that creates riddles about concepts represented as common nouns has been developed. The base of these riddles are comparisons

between the target concept (i.e. *a shirt*) and other entities that share some of its properties (i.e. *snow*).

Later, we explore the relationship between word associations found with already available semantic resources and the type of rhetorical figures that people employ regularly. The aim is to identify possible regularities, intuitions or heuristics that may at a later stage be employed to drive a text generator that can use this type of rhetorical figure. The potential of this system will be studied for the automatic generation of analogies, similes and metaphors.

2.3. Document structure

This document follows a similar structure as the development followed during the project.

First, in Chapter 3, the objectives of this project are proposed and what are the guidelines for obtaining them.

Chapter 4 provides a review of the state of the art both puzzles generation and figurative language generation. An approach of the technologies used for system development and related projects is done.

Once we know the context of language generation and the technologies, Chapter 5 explain the first goal in the development of this project, introducing the process of word association carried out with the Thesaurus Rex tool.

Chapter 6 shows our approach to generate puzzles using word associations. It exposes a first evaluation and the results obtained from a series of questionnaires where people tried to solve two sets of riddles generated by our system. It also contains some conclusions and future work on the generation of riddles, as well as the implementation of some of them.

Afterwards, Chapter 7 shows a second evaluation for the process of word association, but this time is used for rhetorical figures generation. Both the design and the results are detailed, the conclusions obtained and the future work that could be carried out on this application are shown.

Finally, in Chapters 8 and 9 we expose the conclusions obtained as well as the features or modifications that could be added in future iterations of the project.

Capítulo 3

Objetivos

El objetivo principal de este proyecto ha sido desarrollar un generador de recursos lingüísticos y evaluar los distintos resultados para comprobar de qué manera influyen las asociaciones de palabras en el lenguaje natural. En vista de las posibilidades que nos brinda hoy en día la disponibilidad de conocimiento relacionado en la web y a través de diversas herramientas, se ha querido implementar una herramienta que encuentre asociaciones entre conceptos para posteriormente ser utilizadas en adivinanzas y en figuras retóricas.

A la hora de elegir una fuente de conocimiento para establecer el dominio o conjunto de definiciones de conceptos que nuestra herramienta necesita para funcionar, teníamos dos posibilidades: encontrar aplicaciones que gestionasen la información adecuadamente o crear un sistema propio que estructure dicha información y mantenerlo continuamente actualizado. Como ya se ha comentado, era habitual para el desarrollo de este tipo de sistemas el uso de mapas conceptuales u ontologías. Esto se debía a que era necesario el uso de este tipo de dominios para relacionar información y, de este modo, las herramientas pudiesen servirse de ella de forma organizada y estructurada. Estos sistemas requieren mucho tiempo de desarrollo con el consiguiente peligro de no llegar a completarlos y, además, requieren una posterior actualización y mantenimiento de la información.

Un mapa o red conceptual (Wikipedia, 2017a) es un diagrama que representa relaciones sugeridas entre conceptos y es una herramienta gráfica utilizada para organizar y estructurar el conocimiento. Normalmente, consiste en una estructura jerárquica de ramificación descendente donde ideas e información son representadas en forma de cajas o círculos conectados con flechas etiquetadas. Los mapas conceptuales han servido para definir ontologías de sistemas informáticos, como por ejemplo el formalismo del lenguaje de modelado unificado. Los mapas conceptuales se utilizan para estimular la generación de ideas, y se cree que ayudan a la creatividad. El mapeo de conceptos también puede ser visto como un primer paso

en la construcción de ontologías, y se puede usar para representar el argumento formal como en los mapas de argumentos.

Una ontología, en la informática (Wikipedia, 2017b), es una denominación formal y definición de tipos, propiedades y relaciones de entidades, ideas y acontecimientos que existen para un dominio particular del discurso y según un sistema de categorías. Por lo tanto, es una aplicación práctica de la ontología filosófica con una taxonomía. Como una ontología puede aplicarse a la resolución de problemas, campos de inteligencia artificial, ingeniería de software, informática biomédica, biblioteconomía, etc. crean ontologías para limitar la complejidad y organizar la información.

También se ha hablado del desarrollo, en la actualidad, de herramientas relacionadas con la Web Semántica. En este ámbito, también tienen cabida los corpus lingüísticos de texto, los cuales realizan un estudio del lenguaje expresado en *corpora* (muestras) de texto del "mundo real". Consta de un conjunto grande y estructurado de textos y se suelen utilizar para realizar análisis estadísticos y pruebas de hipótesis, verificar incidencias o validar reglas lingüísticas dentro de un territorio lingüístico específico. Este método es un enfoque para derivar un sistema de reglas abstractas a partir de un texto, para conocer cómo funciona un lenguaje natural y cómo se relaciona con otro idioma.

Por otra parte, la idea de la Web Semántica se basa en añadir metadatos semánticos y ontológicos de manera formal al contenido, el significado y la relación de los datos de Internet. De esta manera, es posible procesarlos automáticamente por máquinas, permitiendo compartir y utilizar la información de la comunidad. Es una extensión de la Web a través de estándares del World Wide Web Consortium (W3C), estándares que promueven formatos de datos comunes y protocolos de intercambio en la Web, fundamentalmente el Resource Description Framework (RDF). El objetivo es proporcionar un marco común que permite mejorar Internet ampliando la interoperabilidad entre los sistemas informáticos usando agentes inteligentes, programas que buscan información sin operadores humanos.

Como deseamos evitar el desarrollo de un mapa conceptual y queremos contar con grandes cantidades de información para nuestro sistema, debemos buscar una herramienta relacionada con la Web Semántica. Por lo tanto, la primera necesidad que nos encontramos ha sido hallar una herramienta que sirva para nuestros propósitos y que no haya que mantener actualizada. A partir de esta herramienta, deseamos desarrollar un proceso de asociación de palabras a nivel semántico que se pueda utilizar con otros fines como por ejemplo la generación de adivinanzas.

3.1. Científicos

Nuestros objetivos científicos tienen como idea principal el estudio de la asociación de palabras a nivel semántico. A partir de esta idea, se quiere comprobar si es útil para generar recursos lingüísticos. Primero deseamos probar con la generación de adivinanzas y realizar una evaluación, a través de la cual obtendremos el porcentaje de aciertos de dichas adivinanzas. Queremos realizar una evaluación con varias fases y con estos porcentajes sacaremos algunas conclusiones sobre las diferentes fases, como por ejemplo en cuál de ellas se producen más y menos aciertos.

Posteriormente vamos a utilizar el proceso de asociación de palabras para generar figuras retóricas, concretamente metáforas, símiles y analogías. A su vez, queremos realizar una evaluación sobre dichas figuras y compararlas con dos conjuntos más para establecer líneas base diferentes, un conjunto estará basado en figuras generadas por humanos ya comúnmente aceptadas y otro en conceptos aleatorios elaborado manualmente. Intentaremos sacar conclusiones sobre cómo los humanos evaluamos las figuras literarias y si se pueden establecer patrones para obtener mejores resultados. Los objetivos científicos específicos son los siguientes:

- Conocer el estado actual de las herramientas desarrolladas en el campo de la creatividad computacional y la Generación de Lenguaje Natural.
- Encontrar diferentes recursos lingüísticos en línea para extraer información que enriquezca nuestro sistema.
- Posteriormente a la realización de las evaluaciones de nuestro generador de recursos, obtener conclusiones acerca de la utilidad de las asociaciones de conceptos para dicho propósito.

3.2. Tecnológicos

Inicialmente, debemos estudiar las plataformas disponibles y elegir la que más se adecue a nuestras necesidades. También estudiaremos las herramientas lingüísticas existentes en línea o a través de bibliotecas para incluir en nuestro generador de recursos y que aporten el tipo de conocimiento y utilidad que buscamos. Queremos incluir variedad de herramientas lingüísticas para enriquecer y personalizar la aplicación en mayor medida. Por lo tanto, hay que diseñar una arquitectura que combine correctamente las tecnologías elegidas y preparada a posibles cambios, puesto que podemos desear mejorar la aplicación ante nuevos objetivos o necesidades. Dentro de los objetivos tecnológicos, a su vez figura el implementar un servicio web que ponga las funcionalidades de nuestro sistema a disposición de la

comunidad. Este generador de recursos pretende alcanzar los siguientes objetivos tecnológicos:

- Diseñar e implementar un generador de recursos lingüísticos que utilice la asociación de conceptos con una arquitectura que agrupe las tecnologías reunidas con un buen rendimiento. Debe permitir que las consultas sean personalizables y flexibles, para así comparar las distintas salidas y obtener mejores conclusiones.
- Realizar una evaluación de nuestro generador con personas. A partir de la evaluación y de la discusión de los resultados obtenidos, implementar las mejoras oportunas en nuestro generador de recursos.
- Desarrollar un servicio web en línea en algún servidor del departamento disponible para todos los usuarios de Internet, proporcionando la ayuda necesaria para utilizar las funcionalidades que ofrezca el generador de recursos.

3.3. Método de trabajo

Para organizar de forma eficaz el desarrollo de la aplicación y el estudio de los conocimientos necesarios para realizar este trabajo, se propone un planteamiento de la solución para los objetivos anteriormente mencionados.

En primer lugar, debemos estudiar el estado de la cuestión del ámbito en el que se sitúa nuestro proyecto, es decir, se investigarán los artículos publicados y las herramientas desarrolladas en el marco de la Generación de Lenguaje Natural y la creatividad computacional, informándonos de los avances en dicho campo. Comprobaremos si existen trabajos que hayan investigado previamente las asociaciones de conceptos de la forma en que estamos interesados.

Después de informarnos sobre el estado actual y decidir aquello que utilizaremos en relación con nuestro trabajo, proseguiremos con la labor de investigación. En este caso, debemos estudiar los sistemas existentes relacionados con recursos lingüísticos en línea, y encontrar herramientas útiles para enriquecer la aplicación que deseamos desarrollar, después de analizar en detalle nuestros requisitos y necesidades. Decidiremos qué herramientas incluir como fuentes de conocimiento en nuestra aplicación después de estudiar las diferentes ventajas e inconvenientes de cada una de ellas.

El tercer paso es desarrollar una aplicación que cumpla los objetivos tecnológicos fijados. Como se desea desarrollar una herramienta que posteriormente pueda ser actualizada, se implementará con un entorno de desarrollo en el lenguaje de programación actual como es Java. Debemos implementar una herramienta que reciba

los resultados de consultas a aplicaciones ya disponibles en línea y los interprete correctamente almacenándolos en estructuras de información internas. A partir de esta información disponible en nuestro formato interno, podremos desarrollar las funcionalidades que nos interesen de forma independiente al recurso externo.

Habiendo implementado el generador de recursos lingüísticos, con las funcionalidades necesarias desarrolladas, generaremos unos ejemplos para realizar una evaluación de los resultados del generador. Llevaremos a cabo la evaluación a través de alguna plataforma en línea y analizaremos los resultados obtenidos. Una discusión de los mismos nos ayudará a entender la visión de los recursos lingüísticos por parte de las personas que evaluarán estos conjuntos.

Por último, posteriormente a la discusión de los resultados, obtendremos conclusiones que, por una parte, nos servirán a nivel científico para analizar la utilidad de las asociaciones de conceptos en la generación de recursos lingüísticos y, por otra parte, nos ayudarán a realizar mejoras y actualizaciones sobre nuestra aplicación.

Capítulo 4

Estado de la cuestión

Es necesario un estudio de los avances más actuales en el ámbito de la creatividad computacional y la Generación de Lenguaje Natural. A continuación, se presentan los trabajos previos relacionados con la generación de recursos lingüísticos y algunos conceptos necesarios para la comprensión del sistema que se ha desarrollado.

4.1. Conceptos lingüísticos: Figuras retóricas

Las figuras retóricas o del discurso se presentan cuando se utilizan las palabras de forma no convencional, y al alejarse de ese uso habitual, resultan expresivas. Por esta razón, su uso es propio de las obras literarias.

Un tropo literario (Wikipedia, 2017c) es el uso de lenguaje figurado con fines artísticos, y dentro de las figuras retóricas, son aquellas que cambian el significado general de las palabras. Deriva de la palabra griega *tropos*, que significa cambio o giro, lo que implica que el término se usa para denotar, entre otros, el lenguaje metafórico. El lenguaje figurado es aquel que se opone al lenguaje literal, distintivo de algunos campos del análisis del lenguaje, como la retórica o la semántica. Mientras que en el lenguaje figurado, una palabra expresa una idea en términos de otra, apelando a una semejanza real o imaginaria, en el lenguaje literal se supone que las palabras tienen el sentido que define su significado exacto.

Un tropo literario es la sustitución de una expresión por otra (palabra, frase, o incluso una imagen) cuyo sentido es figurado. Estos ocupan un importante lugar en el lenguaje literario, más concretamente en la lírica. Algunos de los tropos más habituales son la metáfora, la ironía, la alegoría, la hipérbole, la metonimia o la sinécdoque, pero nosotros nos vamos a centrar en la metáfora, el símil y la analogía, dado que son las que nos han interesado en este proyecto. Nos interesan por ser

las figuras que más juego dan a la hora de utilizar asociaciones de conceptos, dado que implican un desplazamiento de propiedades entre dos conceptos.

- Una *metáfora*: Su estudio se remonta a Aristóteles y su Retórica. Una metáfora se refiere a una cosa mencionando otra, utiliza el desplazamiento de características similares entre dos conceptos con fines estéticos o retóricos. Por ejemplo, de una persona valiente se dice que tiene corazón de león. Por lo tanto, la metáfora es una asociación de elementos con similitudes compartidas de modo que se puede sustituir el uno por el otro. Algunas figuras literarias como la antítesis, la hipérbole, la metonimia, el símil o la analogía pueden clasificarse como tipos de metáfora. La metáfora también se puede describir en dos partes, tenor y vehículo. El tenor es el sujeto al cual se atribuyen los atributos. En el ejemplo anterior sería la persona que es valiente y que recibe las cualidades del león. El vehículo es el objeto cuyos atributos se prestan, que sería el corazón de león. La lingüística cognitiva utiliza los términos destino y fuente, para denotar tenor y vehículo.
- Un *símil*: Realiza una comparación entre términos. Un símil es una metáfora dirigida, directa y explícita en la que se comparan dos entidades diferentes con el objetivo de desarrollar un nuevo significado. A pesar de que los símiles y las metáforas son similares, los símiles utilizan explícitamente, aunque no necesariamente, conectores (por ejemplo, como, cual, que, o varios verbos tales como se asemejan). Son muy utilizados en formas de poesía que comparan lo inanimado y lo vivo, como por ejemplo labios como espadas.
- Una *analogía*: Es la comparación entre varios conceptos, indicando las características que permiten dicha relación. En la retórica, una analogía es una comparación textual que resalta alguna de las similitudes semánticas entre los conceptos protagonistas en dicha comparación. Por ejemplo, sus manos son suaves como el terciopelo o sus dientes son blancos como perlas.

4.2. Herramientas de asociación de conceptos

Con el fin de reunir información acerca de las propiedades que caracterizan el concepto destino de la asociación, y para obtener conceptos similares de acuerdo con esas características, hemos utilizado un recurso de asociación de palabras llamado Thesaurus Rex. Por lo tanto, para entender cómo funciona este sistema, primero hay que entender cómo funciona Thesaurus Rex. Además, se presentan otras herramientas también utilizadas para la generación tanto de adivinanzas como de figuras retóricas.

4.2.1. Thesaurus Rex

Thesaurus Rex (Veale y Li, 2013) es un servicio web¹ que dados dos conceptos (por ejemplo, war y divorce) devuelve una nube de las categorías matizadas compartidas por ambos conceptos (en el ejemplo dado devuelve una nube que contiene *traumatic-event*, *stressful-event*, *unexpected-event*...).

Thesaurus Rex organiza conceptos de acuerdo a unas determinadas categorías de grano fino, utilizadas por los hablantes en el lenguaje cotidiano, como por ejemplo *food*, *drink*, *beverage*, etc. Estas categorías tienen un peso asociado que representa su importancia relativa respecto a un concepto dado. Thesaurus Rex puede mostrar diferentes categorías para cada concepto y permite a su vez consultar los conceptos de cada categoría, tal y como muestra la Figura 4.1 para el concepto *coffee*.

Si tomamos como ejemplo el concepto *coffee*, tenemos que algunas de sus categorías con mayor peso son *beverage* o *drink* y algunas de las de menor peso son *seed* o *poison*. La Tabla 4.1 muestra algunas categorías de *coffee* y sus correspondientes pesos. Para el mismo concepto *coffee*, los atributos con mayor peso son *hot* y *acidic*, mientras que los de menor peso son *granulated* y *digestive*, como se muestra en la Tabla 4.2.

| Categoría | Peso |
|------------|------|
| drink | 4983 |
| smell | 185 |
| beverage | 7056 |
| seed | 3 |
| intoxicant | 14 |
| liquid | 2541 |
| food | 3322 |
| poison | 5 |
| ... | ... |

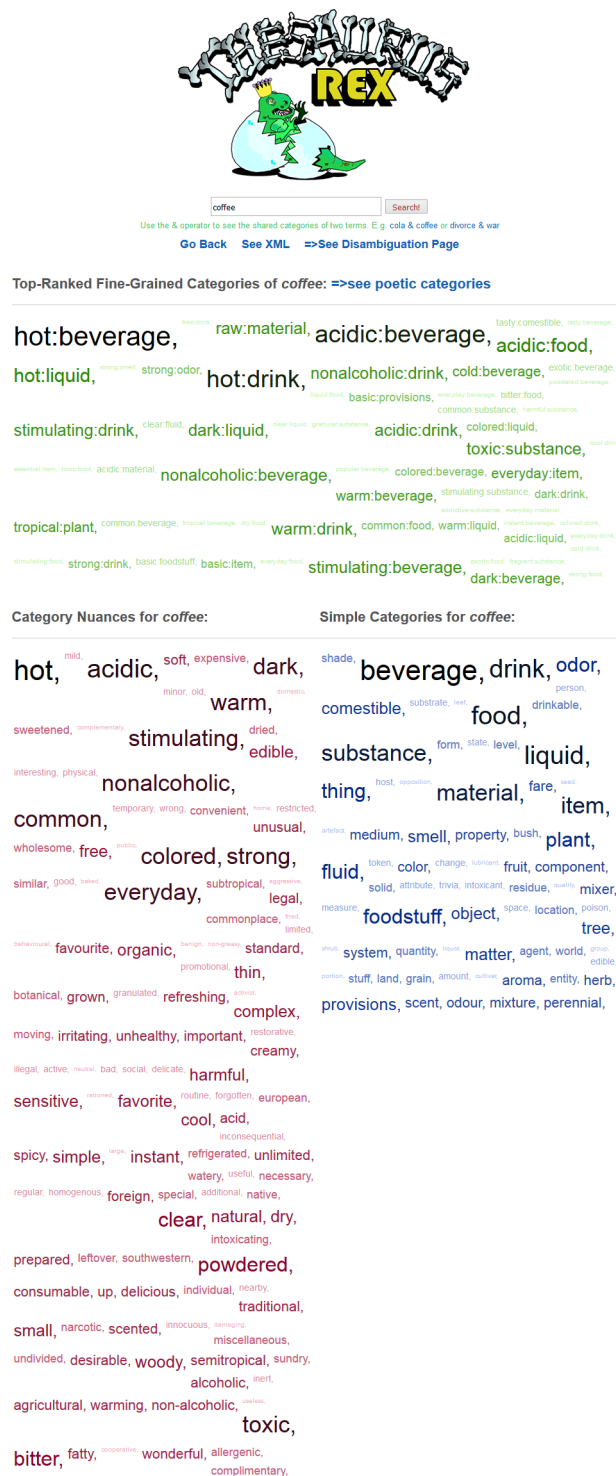
Tabla 4.1: Ejemplos de categorías (con sus pesos) para el concepto *coffee* en Thesaurus Rex

| Atributo | Peso |
|---------------|------|
| hot | 3900 |
| granulated | 10 |
| acidic | 2909 |
| dark | 1144 |
| stimulating | 1267 |
| noncarbonated | 24 |
| colored | 696 |
| digestive | 7 |
| ... | ... |

Tabla 4.2: Ejemplos de propiedades (con sus pesos) para el concepto *coffee* en Thesaurus Rex

Thesaurus Rex utiliza la Web Semántica para generar sus resultados, con lo cual la información disponible no es fija, sino que varía según los datos actuales de la web. La ventaja de utilizar esta herramienta es que se encuentra en continua actualización, pero el inconveniente es que en algunos casos la información puede

¹<http://ngrams.ucd.ie/therex2/>

Figura 4.1: Captura del servicio web Thesaurus Rex para el concepto *coffee*

resultar un poco extraña dado que se crea semiautomáticamente desde contenido de la web. El generador de acertijos que proponemos recibe un nombre común de entrada, que es el concepto final del acertijo. Usando esta base de datos de asociaciones de palabras extraídas de la web, nuestro sistema despliega una serie de comparaciones entre el concepto destino y otros conceptos con propiedades similares con el fin de crear la adivinanza final.

4.2.2. ConceptNet

ConceptNet² es una red semántica multilingüe gratuita de código abierto, creada para ayudar a entender significados de conceptos en lenguaje natural. Nació dentro del proyecto *crowdsourcing* Open Mind Common Sense (MIT Media Lab, 1999). Inicialmente se basaba únicamente en la información de la base de datos OMCS, para más adelante incluir conocimiento de variedad de recursos, como otros recursos *crowdsourced* o de colaboración abierta distribuida (como Wiktionary además de Open Mind Common Sense), juegos con un propósito (como Verbosity y nadya.jp) y recursos creados por expertos (por ejemplo, WordNet y JMDict).

ConceptNet queda reflejado como un grafo dirigido cuyos nodos representan conceptos estrechamente relacionados, los cuales son conjuntos de frases en lenguaje natural tales como frases nominales, verbales, adjetivas, etc. Las relaciones que existen entre estos conceptos, relaciones entre pares de conceptos, son representadas por las aristas que conectan los nodos del grafo, las cuales tienen asignado un peso que indica lo importante e informativa que es dicha relación. Estas aristas son relaciones habituales de sentido común dentro de un conjunto limitado de posibles relaciones y representan patrones de oraciones comunes que se encuentran en el corpus de OMCS. Algunas de ellas son IsA, Has, IsUsedFor, PartOf, MadeOf, SimilarTo o HasProperty.

4.2.3. WordNet

WordNet es una base de datos léxica, de palabras en inglés unidas entre sí por sus relaciones semánticas. Incluye las categorías léxicas de sustantivos, verbos, adjetivos y adverbios, ignorando preposiciones, determinantes y otras palabras funcionales. Las palabras se agrupan en conjuntos no ordenados de sinónimos (syn-sets), cada uno de los cuales expresa un concepto distinto. Proporciona definiciones cortas y ejemplos de uso de dichos conjuntos, y registra un número de relaciones conceptuales-semánticas y léxicas entre ellos o sus miembros.

WordNet se asemeja a un diccionario de sinónimos sobrealimentado con una estructura de grafo que agrupa las palabras en base a sus significados. No sólo enlaza

²<http://conceptnet5.media.mit.edu/>

las formas de las palabras, sino también sus significados, de forma que las palabras que se encuentran próximas entre sí en la red son desambiguadas semánticamente. Además, WordNet etiqueta las relaciones semánticas entre palabras, al contrario que un diccionario de sinónimos, donde los grupos de palabras no siguen ningún patrón que no sea la similitud de significado.

WordNet posee una interfaz en línea³ accesible por cualquier usuario, donde se puede navegar por la red resultante de palabras y conceptos relacionados, pero su principal uso es en el análisis automático de texto y aplicaciones de inteligencia artificial. La base de datos y herramientas de software están publicadas bajo una licencia BSD y disponibles para descargar gratuitamente.

4.3. Generación de acertijos

Aunque la generación de acertijos puede parecer una tarea difícil desde el punto de vista computacional, ha habido intentos previos de generación automática de acertijos. Dentro de estos sistemas, se incluyen generadores de adivinanzas en forma de pregunta sencilla con respuesta sencilla, juegos de palabras, pregunta en forma de analogía, siguiente en la secuencia o preguntas acerca de personajes y celebridades.

4.3.1. Adivinanzas: accesibilidad y representación de la información

De Palma y Weiner (1993) determinaron que las jerarquías de accesibilidad entre los significados de las palabras polisémicas juegan un importante papel en la generación y la comprensión de una categoría de acertijos sencillos. En otro estudio, De Palma y Weiner (1992) proponen un modelo de representación del conocimiento que contiene los datos para generar o resolver acertijos. Han desarrollado un algoritmo que genera una suposición o conjetura basada en conceptos homófonos. Pero restringen su atención a acertijos que se ajustan al siguiente formato, donde el acertijo consiste en una pregunta sencilla seguida de una sencilla respuesta:

Pregunta: What has a mouth and cannot eat?

Respuesta: A river

4.3.2. JAPE

JAPE (Binsted y Ritchie, 1997; Ritchie, 2003) es un programa de ordenador que genera acertijos con juegos de palabras simples usando plantillas con espacios donde se insertan las palabras o frases. Para determinar qué palabras deben ser

³<http://wordnetweb.princeton.edu/perl/webwn>

incorporadas a la adivinanza final, el sistema hace uso de esquemas predefinidos (manualmente contruidos a partir de chistes conocidos previamente). Estos esquemas establecen relaciones entre las palabras que se deben mantener para construir el juego de palabras.

El programa fue probado por 120 niños que puntuaron la comicidad tanto de las adivinanzas generadas, como de textos generados por humanos, y textos no cómicos. La evaluación confirmó que las adivinanzas generadas eran graciosas, y que no hay diferencia significativa en la comicidad entre los acertijos con juegos de palabras generados por su sistema y aquellos publicados que han sido generados por humanos.

Un ejemplo de juego de palabras generado por JAPE es el siguiente:

What is the difference between leaves and a car?

One you brush and rake, the other you rush and brake

Posteriormente, algunos de los autores de JAPE (Cunningham et al., 2000) han desarrollado STANDUP (Waller et al., 2009), un generador de juegos de palabras a gran escala para permitir que los niños con discapacidades de comunicación mejoren sus habilidades lingüísticas. Con el fin de adaptar los juegos de palabras generadas al público final (es decir, los niños con discapacidades de comunicación), además de la generación de un juego de palabras a través de los mismos pasos utilizados en JAPE, tuvieron que introducir varias mejoras como salida de voz, soporte de imagen, temas restringidos o uso de palabras familiares. El sistema se evaluó con usuarios reales en un período corto, y aunque no se pudieron observar efectos positivos a largo plazo, los autores informan de un cambio en la actitud de los niños hacia la comunicación.

4.3.3. Generación automática de acertijos

Colton (2002b) primero desarrolló el programa HR, que implementa un modelo para la creación de teorías basado en un ciclo de actividad matemática o dominios matemáticos, por el cual, se forman conceptos, conjeturas acerca de estos conceptos e intentos de resolver las conjeturas. HR tiene diversas medidas que estiman el grado de interés de un concepto. Utilizando el generador de modelo MACE (McCune, 1994a) genera objetos de interés a partir de conjuntos de axiomas y realiza la formación de conceptos. Durante la formación de conceptos, HR utiliza pruebas empíricas que sugieren conjeturas y emplea el demostrador de teoremas Otter (McCune, 1994b) para probar una conjetura dada.

Más tarde, Colton (2002a) proporcionó una caracterización de ciertos tipos de adivinanzas según la estructura de la pregunta planteada y la respuesta a las mismas. Con esta caracterización, amplió el sistema HR para generar adivinanzas

automáticamente dada información sobre un conjunto de objetos de interés. La principal dificultad técnica fue garantizar la unicidad y simplicidad del concepto que explica la solución del acertijo generado. Se generan tres tipos de preguntas:

- **Extraño o fuera de lugar:**

P. Which is the odd one out: hair, triangles, squares, plants, words, trees?

R. Triangles: all the others have roots

- **Analogía:**

P. Hair is to stubble as potatoes are to: (a) french fries (b) sweet potatoes (c) potato skins (d) vegetable?

R. French fries

- **Siguiente en la secuencia:**

P. What is next in the sequence: 3, 8, 15, 24, 35?

R. 48: Starting from 2, square each consecutive integer then subtract 1

4.3.4. Utilizando modelos temáticos y medidas de similitud semántica

Pintér et al. (2012) proponen un modelo basado en el conocimiento para generar tres tipos de acertijos (fuera de lugar, elegir la palabra relacionada y separar los temas) a partir de colecciones de documentos no estructuradas y no anotadas. Los acertijos generados se pueden ajustar a dos niveles de dificultad: principiante e intermedio. El algoritmo se basa en modelos temáticos, similitud semántica, y capacidad de la red. Este método pretende ayudar a los diseñadores de adivinanzas a recopilar una colección de juegos de palabras de forma semiautomática.

Un ejemplo de fuera de lugar (*odd one out*) sería:

Pregunta: Salmon, shark, whale, elephant?

Respuesta: Elephant, because all the others live in water

4.3.5. TheRiddlerBot

Guerrero et al. (2015) presentan un bot de Twitter (@TheRiddlerBot⁴) que genera acertijos sobre celebridades, presenta sus salidas como concursos a sus seguidores, comprueba las respuestas publicadas y responde en consecuencia. El modelo selecciona una celebridad, recupera rasgos relevantes para describirlo, genera analogías entre sus atributos y convierte dichas descripciones en enunciados. Como se puede observar en la Figura 4.2, para el ejemplo de *Forrest Gump* se indica que es una persona desagradable, tonta pero cariñosa, vive en Estados Unidos y

⁴<https://twitter.com/theriddlerbot>

es el Homer Simpson de Bubba Gump. Después de este proceso, el bot muestra el enigma generado en Twitter e interactúa con los usuarios mediante la evaluación de sus respuestas.



Figura 4.2: Captura de la página de TheRiddlerBot en Twitter

Para evaluar la generación de adivinanzas, 86 personas valoraron cinco acertijos. Primero se pidió a los participantes adivinar la respuesta al enigma. Entonces, presentaban la respuesta correcta y preguntaban si conocían a la persona en cuestión. El porcentaje de celebridades conocidas fue 54,19 %. El bajo número de respuestas correctas (15,58 %) sugirió que la complejidad de los acertijos generados era alta. Este es un ejemplo de acertijo generado por su sistema:

Tell me the name of a person that is the Morpheus of The Dark Knight Rises, is criminal, playful yet cruel, has been seen wearing a purple topcoat. Who is this?

4.4. Generación de figuras retóricas

Las figuras retóricas (véase la sección 4.1) juegan un importante papel en la comunicación, ya que se producen con una frecuencia de una por cada tres frases (Shutova et al., 2012), por lo que la generación de este tipo de recurso lingüístico es esencial para la Generación de Lenguaje Natural.

Black (1955) hace explícito que las metáforas dependen de conexiones conceptuales entre redes o mapas conceptuales. La idea de que las metáforas son una cuestión de mapeo de dominios cruzados (Lakoff, 1993) es inherente a este enfoque. Una metáfora es un proceso cognitivo que construye o mapea las conexiones entre redes conceptuales, como ocurre con los símiles y las analogías. En consecuencia, para generar estas figuras literarias se necesita una estructura conceptual donde se

coloca cada concepto no sólo teniendo en cuenta su uso convencional, sino también sus usos diversos y no convencionales (Veale, 2014).

El mejor lugar para encontrar esta estructura compleja es la Web y es donde vamos a buscar asociaciones de palabras con el fin de crear nuestras analogías, símiles y metáforas.

Las figuras retóricas han sido objetivo de investigadores en métodos computacionales de la lingüística de forma intermitente durante muchos años. Sin embargo, sólo en los últimos años existe la combinación de recursos de conocimientos disponibles y conocimientos acumulados que permiten que este campo florezca.

Las metáforas han sido extensamente estudiadas en Análisis de Lenguaje Natural pero no tanto en Generación de Lenguaje Natural (GLN). Hay una gran cantidad de trabajos relacionados con la detección de metáforas (Wilks et al., 2013), la identificación de metáforas (Shutova et al., 2010), extracción y anotación de metáforas (Wallington et al., 2003) pero pocos relacionados con la generación de metáforas. Puede que la razón sea que la generación de metáforas es tan difícil como la creatividad humana se lo permite. En esta sección se presentan los enfoques más importantes para la generación de metáforas, símiles y analogías.

4.4.1. Aproximaciones a las figuras retóricas en GLN

En el campo de la Generación de Lenguaje Natural se han producido una serie de intentos de establecer procedimientos para la construcción de figuras retóricas como ingredientes importantes de fragmentos de texto generados. Esto se ha intentado tanto en términos generales (Hervás et al., 2006b) para diferentes tipos de figuras retóricas, y para casos específicos como analogías (Hervás et al., 2006a) o metáforas (Hervás et al., 2007). Estos intentos se realizaron antes de que estuviesen disponibles fuentes adecuadas de conocimiento legible por máquinas y, en consecuencia, sufrieron de la escasez de conocimientos apropiados. En estos intentos se consideró el problema del emplear figuras retóricas en la generación de texto en términos generales, pero carecían de suficiente cantidad de conocimiento de la semántica de las palabras para ser capaces de una generación práctica.

4.4.2. Aproximaciones para la construcción conceptual de figuras retóricas

El reciente desarrollo de fuentes de conocimiento que permiten una fácil extracción de grandes corpus de texto para asociaciones de palabras significativas ha dado lugar a la aparición de una serie de sistemas que dependen de éstos para la construcción de figuras retóricas de diferentes tipos.

Para entender bien el mecanismo de dichos sistemas conviene conocer el concep-

to de n-grama dentro de la lingüística computacional. Un n-grama es una secuencia contigua de n elementos (que pueden ser fonemas, sílabas, letras, palabras, etc.) a partir de una secuencia de texto o de discurso. Un modelo de n-grama es un tipo de modelo probabilístico que permite hacer una predicción estadística del próximo elemento de cierta secuencia de elementos sucedida hasta el momento y sus principales ventajas son simplicidad y escalabilidad. Originalmente se utilizó (Shannon, 1951) para estimar ratios de entropía y redundancia en el Inglés escrito y en la llamada teoría matemática de la información (Shannon, 1948). Shannon planteó que dada una secuencia de letras (por ejemplo, la secuencia "for ex"), ¿cuál es la siguiente letra con mayor probabilidad? Desde entonces, se han analizado los n-gramas más frecuentes y algunas empresas como Google poseen gran cantidad de este tipo de información obtenida de analizar más de cinco millones de libros. A través de Google Ngram Viewer⁵ se puede acceder a la información de Google n-gram corpus y aprovecharla para implementar sistemas como reconocimiento de voz o recomendaciones en editores de texto.

4.4.3. Jigsaw Bard

Jigsaw Bard (Veale y Hao, 2011) es un servicio web⁶ que explota esta noción de readymades o confecciones lingüísticas para generar nuevos símiles creativos bajo demanda. Jigsaw Bard escanea n-gramas de Google para indexar un gran número de confecciones lingüísticas potenciales que puedan en un futuro ser reutilizadas como símil creativo por alguna solicitud. Siguiendo los principios de la arquitectura orientada a servicios, Jigsaw Bard hace el mínimo procesamiento en tiempo real que sea posible. De modo que, cuando es llamado como un servicio web, recupera de forma fiable decenas de símiles que ya habían sido indexados para una propiedad.

Por ejemplo, dada la propiedad adjetiva *quiet*, Jigsaw Bard devuelve el símil "*The peaceful life of a monastery*". Jigsaw Bard se entiende mejor como un diccionario de sinónimos creativo: para cualquier propiedad dada (o combinación de propiedades) seleccionada por el usuario, presenta una serie de símiles aptos, y los usuarios deben decidir qué símiles son los más adecuados para sus fines descriptivos.

4.4.4. Thesaurus Rex

Thesaurus Rex (Veale y Li, 2013) utiliza la web para generar una amplia lista de perspectivas alternativas sobre ideas familiares, y selecciona algunas de ellas con WordNet para realizar un razonamiento vertical y producir juicios de similitud. Thesaurus Rex también utiliza la perspectiva con mayor información para

⁵<https://books.google.com/ngrams>

⁶<http://afflatus.ucd.ie/jigsaw/>

explicar de forma concisa cada comparación, o sugerir una comparación creativa. Por ejemplo, para acentuar la toxicidad potencial del café, Thesaurus Rex sugiere comparaciones con alcohol, tabaco o pesticida, ya que todos han sido clasificados como sustancias tóxicas en la web.

Este sistema puede ver *champán* como algo más que un alimento, “*artículos de lujo como diamantes y champán*” como indica un fragmento de la web. Thesaurus Rex está formado por estas perspectivas de grano fino superpuestas, combinadas en un servicio web para la categorización, y de esta forma, proporciona unas posibilidades que WordNet no posee. A Thesaurus Rex se le puede preguntar qué tienen en común los conceptos lexicalizados *war & peace*, *life & death* o *divorce & war*. La Figura 4.3 muestra una nube con las categorías de matices compartidas por *divorce* y *war*, entre los que se encuentran *traumatic-event*, *stressful-event*, *unexpected-event*. . .

Figura 4.3: Captura de la aplicación Thesaurus Rex, para la búsqueda de conceptos *divorce and war*

Thesaurus Rex organiza conceptos de acuerdo a categorías utilizadas por los hablantes en el lenguaje cotidiano (*food, drink, beverage*. . .). Estas categorías tienen asociado un peso que representa su importancia relativa para el concepto dado. Thesaurus Rex puede mostrar diferentes categorías para cada concepto y permite a su vez consultar los conceptos en cada categoría. Por ejemplo, para el concepto *coffee*, algunas de sus categorías con más peso son *beverage* o *drink* y algunas con menos peso son *leaf* o *apposition*. Los conceptos en Thesaurus Rex tienen propiedades o modificadores asociados que van acompañados de un peso no estándar que indica cómo es de fuerte su relación con el concepto. Por ejemplo, para *coffee* algunos de los modificadores con más peso son *hot*, *acidic* o *stimulating*, y algunos modificadores con menos peso son *smaller* o *adult*.

Thesaurus Rex ofrece un valioso servicio a cualquier sistema que desee tener

una visión divergente de conceptos, ya sea con la finalidad de evaluar similitudes de conceptos de forma literal o para el razonamiento no literal en el caso de las metáforas. Rex se puede utilizar como una herramienta de navegación para usuarios web en busca de ideas o comparaciones válidas, o como un servicio flexible de similitudes que soporta sistemas de procesamiento de metáforas.

4.4.5. Metaphor Magnet

Metaphor Magnet (Veale y Li, 2012) es un sistema implementado como servicio web⁷ que genera y expande metáforas afectivas bajo demanda. Permite a los usuarios introducir consultas de términos simples (tales como *leader*), compuestos por términos con un elemento afectivo (como *good leader* o *+leader*), o declaraciones (tales como “*Steve Jobs is a +leader*”). Por cada entrada, para generar las interpretaciones afectivas más adecuadas, el servicio une su conocimiento de los estereotipos lexicalizados con los n-gramas de Google a gran escala. En cada caso, Metaphor Magnet proporciona una explicación de sus salidas a nivel de propiedades. Así, por ejemplo, si *Steve Jobs* tuviera que ser visto como un *master*, las propiedades *skilled*, *enlightened*, *free* y *demanding* serían todas destacadas como las más apropiadas.

En la visión de propiedades teóricas de la metáfora que ofrece Metaphor Magnet entiende la interpretación de metáforas como una cuestión de qué propiedades se asignan desde el vehículo al tenor, de la fuente al destino. Parte de un modelo del mundo totalmente basado en propiedades y en el que las palabras expresan estereotipos que se correlacionan con propiedades sobresalientes. Metaphor Magnet posee una visión a nivel proposicional del mundo, en el que los estereotipos están vinculados a otros estereotipos por relaciones arbitrarias.

4.4.6. Metaphor Eyes

Metaphor Eyes⁸ (Veale, 2014), a diferencia de Metaphor Magnet emplea un modelo proposicional del mundo que funciona con tripletas sujeto-relación-objeto en vez de pares de objeto-atributo.

Metaphor Eyes recupera vehículos aptos para un determinado concepto tenor. Por lo tanto, dado *philosopher* como tenor, Metaphor Eyes considera *scholar*, *moralist*, *theologian*, *historian*, *scientist*, *visionary*, *explorer* e incluso *warrior* como vehículos para una metáfora. Para cualquiera de estos conceptos dados, intenta adaptar su conocimiento sobre él a su representación del tenor, considerando las proposiciones asociadas con el vehículo que pueden convertirse en proposiciones

⁷<http://ngrams.ucd.ie/metaphor-magnet-acl/>

⁸<http://ngrams.ucd.ie/metaphor-eye/>

aptas sobre el tenor. Considerando la opción de *explorer*, se produce la metáfora *philosophers are explorers*. Sabiendo que los exploradores realizan peregrinaciones, Metaphor Eyes busca en los n-gramas de Google una o más proposiciones del mismo tipo, como “*philosopher’s quest*”.

Al considerar la metáfora como una forma de representación, se permite rellenar los huecos en el caso de una comprensión pobre de un concepto mediante la importación de conocimiento de un concepto vecino. Por lo tanto, se ofrece una solución parcial a la metáfora como problema, a la vez que se muestra la metáfora como una solución en sí misma. Metaphor Eyes metaforiza un concepto (la fuente) como otro concepto (el destino). Dados los conceptos *Scientist* y *Artist*, se generan metáforas como “*Scientists develop ideas like artists*”.

4.4.7. FIGURE8

FIGURE8 (Harmon, 2015) es un sistema que contiene un modelo subyacente para el que define comparaciones creativas y figurativas, y evalúa su salida en base a estas reglas. FIGURE8 es un sistema que utiliza un enfoque basado en la web para formar una base de conocimiento preliminar de sustantivos y sus propiedades.

El sistema está provisto con un modelo del mundo actual y una entidad en el mundo que se describirá. Se selecciona un vehículo adecuado de la base de conocimiento, y se explica la comparación entre los dos conceptos. Para aclarar esta comparación, se obtiene un entendimiento de lo que estos conceptos pueden hacer y cómo se pueden describir, a través de una búsqueda en un corpus. Se completan las frases a través de una biblioteca de casos de construcciones gramaticales válidas. Por último, la comparación es clasificada por el sistema en base a cualidades semánticas, fonéticas, y basadas en conocimiento. De esta manera, FIGURE8 simula el proceso de revisión mediante la generación de diversas opciones de vehículos y variaciones lingüísticas para un solo tenor, eligiendo como su favorito al mejor de ellos.

FIGURE8 proporciona una nueva base para un sistema de generación y evaluación de lenguaje figurativo inteligente (descripciones figurativas creativas), a pesar de que no pretende tener un amplio conjunto de reglas. Las evaluaciones de FIGURE8 se basan en modelos psicológicos de comprensión de la metáfora, y se descubrió que se ajustaban mucho a las clasificaciones humanas.

Capítulo 5

Diseño e implementación de un proceso de extracción de relaciones entre conceptos

En este capítulo se presenta el desarrollo del proceso de asociación de conceptos, con sus entradas, su proceso y su salida. Posteriormente, se introduce la información correspondiente al servicio web implementado a partir de dicho proceso de relación de conceptos.

5.1. Proceso de asociación de palabras

El proceso desarrollado encuentra asociaciones de palabras, a través de diferentes consultas al servicio Thesaurus Rex, cuyas respuestas son primero analizadas e interpretadas. Siguiendo los pasos descritos en la siguiente sección 5.1.2, el sistema genera los conceptos fuente con propiedades similares a las del concepto destino creando asociaciones de palabras.

5.1.1. Entrada

El enfoque propuesto recibe un nombre común como entrada, que es el concepto destino para el que se deben generar asociaciones de palabras. Es decir, debemos encontrar conceptos fuente que contengan propiedades que caracterizan al concepto recibido de forma más fuerte o representativa. Utilizando Thesaurus Rex (véase las secciones 4.2.1 y 4.4.4), el sistema crea una comparación entre el concepto destino y el concepto fuente a través de la propiedad que comparten, con el fin de crear una asociación de palabras.

5.1.2. Proceso

La Tabla 5.1 muestra una serie de ejemplos de los conceptos destino y cómo Thesaurus Rex es usado para generar las comparaciones. A continuación, partiendo de *snow* como primer concepto, se muestra en el siguiente ejemplo el proceso detallado para generar un acertijo.

| Paso | Concepto destino | snow | network |
|------|--|--|--|
| 1 | Categorías | surface, weather, elements... | system, entity, structure... |
| 2 | Modificadores | natural, reflective, slippery , soft... | social , complex, adaptive, physical... |
| 3 | Categorías para el modificador seleccionado | surface , ground, stuff... | institution , event, activity, science... |
| 4 | Nueva búsqueda | slippery surface | social institution |
| 5 | Conceptos obtenidos | satin, silk , nylon, polyester... | family, government , religion... |

Tabla 5.1: Ejemplos de asociaciones de palabras obtenidas. Las palabras en negrita representan las elecciones hechas para cada ejemplo.

1. **Categorías del concepto destino.** Para obtener las categorías deseadas a las cuales pertenece el concepto destino, primero se extrae una lista de todas las categorías mediante una búsqueda en Thesaurus Rex. De esta lista, sólo se consideran como candidatas el $L\%$ de las categorías con mayor peso. El valor de $L\%$ es configurable (siendo en este caso $L = 0.4$). Si se establece un valor $M\%$ demasiado alto, tendremos una lista de categorías con pesos bajos, las cuales serán poco relevantes para el concepto destino. De la misma manera, si establecemos un valor $N\%$ demasiado bajo, corremos el riesgo de acortar la lista de modo que contenga tan sólo un elemento. En el ejemplo de *snow*, las categorías con pesos más altos en Thesaurus Rex son *surface* y *weather*.
2. **Selección de modificador**

Extracción del modificador. Además de las categorías, también se necesita una lista de modificadores asociados al concepto destino, que es devuelto por una nueva consulta a Thesaurus Rex. De esta lista, los $P\%$ de los atributos con los pesos más altos se consideran como candidatos (en este ejemplo, $P = 0.6$). Por ejemplo, si nuestro concepto destino es el sustantivo *snow*, al-

gunos de los modificadores extraídos más importantes son: *natural*, *reflective*, *slippery*, *soft* y *white*.

Elección de modificador. Se selecciona uno de los modificadores obtenidos previamente. Esta selección es aleatoria para que el sistema sea menos repetitivo, ya que las palabras asociadas a un mismo concepto destino no son siempre los mismos que si se selecciona sólo el modificador con el peso más alto. Para el ejemplo actual, el sistema ha elegido el modificador *slippery*.

3. Selección de categoría

Extracción de categorías. Partiendo del modificador seleccionado en el paso anterior, se realiza una nueva consulta a Thesaurus Rex, con el fin de obtener las categorías que presentan este modificador como una propiedad representativa. En el ejemplo de *snow*, las categorías seleccionadas podrían ser *surface*, *ground* y *stuff*, que son categorías que presentan la propiedad *slippery* en Thesaurus Rex.

Elección de categoría. Una de las categorías obtenidas en el paso anterior es seleccionada. El sistema puede ser parametrizado para seleccionar una categoría que contenga el concepto destino, como en el ejemplo actual (una categoría que coincida con alguna obtenida en el paso 1). También podría ser parametrizada para elegir una categoría en la que no se incluye el concepto destino (descartando las categorías que coinciden con los obtenidos en el paso 1). Para el ejemplo actual, *surface* es la categoría seleccionada.

4. **Composición de nueva búsqueda.** Entonces se compone una nueva consulta para Thesaurus Rex utilizando la categoría obtenida en el paso anterior y el modificador seleccionado en el paso 3. En el ejemplo actual, esta nueva consulta es *slippery surface*.

5. **Selección de concepto final.** Con la consulta compuesta en el paso anterior, obtenemos una lista de conceptos que pertenecen a la categoría seleccionada en el paso 5 (*surface*) y al mismo tiempo presentes en la propiedad seleccionada en el paso 3 (*slippery*). Esta lista es por lo general bastante amplia, por lo que el sistema elige al azar entre los resultados que tienen un peso asociado entre los R % de los conceptos con los pesos más altos (en este ejemplo, $R = 0,1$). En nuestro ejemplo, los conceptos finales asociados al concepto destino son *satín*, *silk*, *nylon* o *polyester*

En la Figura 5.1 se puede observar el diagrama de flujo que sigue el proceso de obtención del concepto fuente para la asociación de conceptos.

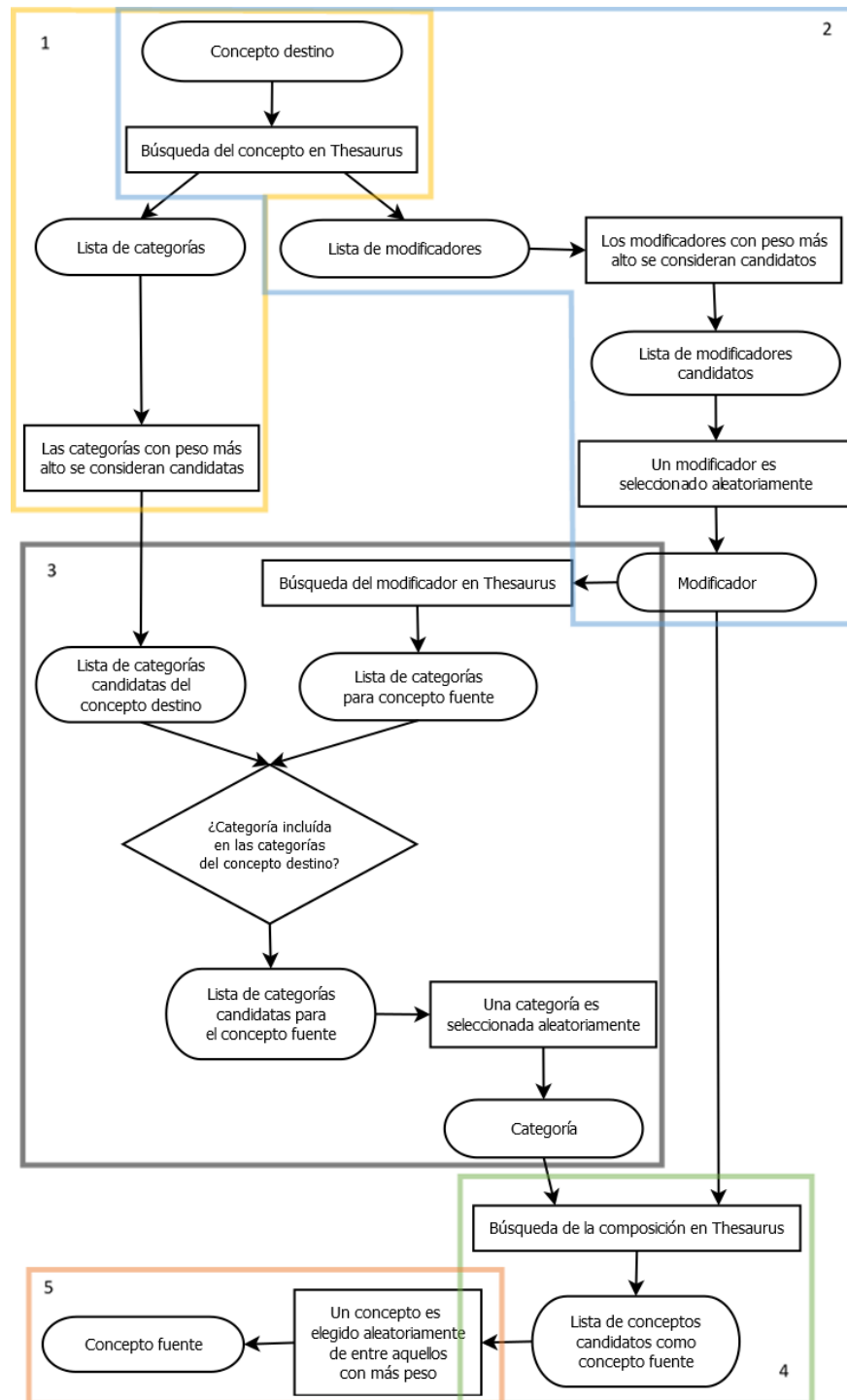


Figura 5.1: Diagrama de flujo del proceso

5.1.3. Salida

La salida del sistema es un concepto relacionado a través de una propiedad compartida con el concepto destino original proporcionado por el usuario. La propiedad compartida es significativa en ambos conceptos, lo que significa que la propiedad tiene un peso alto para ambos. El concepto fuente resultante se elige al azar de la lista de conceptos generados, y se utiliza posteriormente para crear una asociación de palabras.

5.2. Servicio web

Se ha implementado un servicio web RESTful para publicar las funcionalidades de nuestro sistema a través del servidor Hefesto¹, propiedad de la facultad de informática de la Universidad Complutense. Está desarrollado (Peña, 2016) bajo la arquitectura REST (Representational State Transfer), un tipo de arquitectura de interfaces de comunicación basado en un protocolo de cliente/servidor sin estado (protocolo HTTP) para realizar servicios web cacheables y escalables. No es un estándar ni un protocolo, sino una serie de principios de arquitectura con unas operaciones bien definidas en el que los recursos están identificados de forma única por URIs. El protocolo HTTP utilizado en los servicios REST es la base de las comunicaciones en Internet y se basa en una serie de peticiones enviadas por el cliente. Estas peticiones son recibidas y procesadas en el servidor, y devueltas al cliente en forma de respuesta.

Este protocolo no mantiene el estado de las peticiones, es decir, no existe una sesión en la que se almacene el resultado de una petición para determinar futuras peticiones. Como cada petición es totalmente independiente, incluso en aspectos de autenticación, cada petición debe llevar todo el contenido necesario para ser realizadas. Este tipo de protocolos sin estado permite a los servidores ahorrar recursos por no tener que almacenar información de la sesión de usuario y facilita las labores de desarrollo.

Para realizar las peticiones, el protocolo HTTP proporciona una serie de métodos o verbos para las diferentes acciones sobre un recurso. Al acceder desde nuestro navegador a una página web, se realiza una petición a un servidor con el método GET, utilizado para consultas y lectura de los recursos. Sin embargo, si se envía un formulario desde una web, se realiza una petición POST. Los sistemas REST utilizan dichos métodos además de PUT, DELETE, PATCH y algunos más para identificar operaciones sobre los diferentes recursos, pero nuestro sistema solamente necesita métodos GET.

¹<http://hefesto.fdi.ucm.es>

Un recurso debe estar identificado de forma única por una URI, que es una cadena de texto con una estructura determinada que nos permite identificar un elemento. La estructura tiene la siguiente forma:

protocolo://dominio:puerto(opcional)/ruta del recurso?parámetros

La descripción de esta arquitectura no incluye un formato específico de respuesta y los formatos elegidos suelen ser XML o JSON, pues son lenguajes que permiten describir estructuras y jerarquías fácilmente, además de ser extensamente utilizados y conocidos. Nuestro servicio REST es capaz de devolver las respuestas en estos dos formatos. Como las URIs deben ser independientes de dicho formato, el acceso al recurso no debe indicar el formato de acceso. Se indica el tipo de formato en la información adicional de la petición HTTP, en las cabeceras de la petición se especifica `application/xml` o `application/json`. En nuestro caso, la consulta debe ser `http://hefesto.fdi.ucm.es/correlator/riddles?{parámetros}` o `http://hefesto.fdi.ucm.es/correlator/tropes?{parámetros}`.

En la dirección `http://hefesto.fdi.ucm.es/correlator/help?` se encuentra un documento de ayuda para usar el sistema con ejemplos de las posibles consultas. Como contenedor web se ha utilizado Apache Tomcat.

El proyecto se ha implementado bajo el entorno de desarrollo de código abierto multiplataforma Eclipse y se ha convertido posteriormente en un proyecto Maven para ayudarnos con las bibliotecas de dependencia. Se ha desarrollado una aplicación en lenguaje Java de tipo *Dynamic Web Project*. Java define el soporte REST (Wikipedia, 2015) a través de la Solicitud de *Java Specification Request* (JSR) 311, la cual se denomina JAX-RS (The Java API for RESTful Web Services). JAX-RS soporta la creación de XML y JSON a través de la arquitectura Java para XML Binding (JAXB). JAX-RS utiliza anotaciones para definir la relevancia REST de las clases Java, las más importantes son las siguientes:

1. `@PATH(ruta)`: Establece la ruta de acceso a la URL base + `/ruta`. La URL base consta del nombre de la aplicación, el servlet y el patrón URL del archivo de configuración `web.xml`. Por ejemplo `@PATH(/help)` dirige a la página de ayuda de nuestro sistema.
2. `@POST`: Indica que el siguiente método responderá a una solicitud HTTP POST.
3. `@GET`: Indica que el siguiente método responderá a una solicitud HTTP GET.
4. `@PUT`: Indica que el método siguiente responderá a una solicitud HTTP PUT.

5. `@DELETE`: Indica que el siguiente método responderá a una solicitud HTTP DELETE.
6. `@Produces(MediaType.TEXT_PLAIN[, más_tipos])`: Define qué tipo MIME se entrega mediante un método anotado con `@GET`. En nuestro sistema, utilizamos `@Produces(MediaType.APPLICATION_XML, MediaType.APPLICATION_JSON)` para indicar dos tipos de salida, XML y JSON.
7. `@Consumes(type[, más_tipos])`: Define qué tipo MIME es consumido por este método.

Jersey framework es un marco de código abierto para el desarrollo de servicios Web RESTful en Java. Ofrece soporte para las API de JAX-RS y proporciona su propia API que amplía el kit de herramientas JAX-RS. Este marco sirve como implementación de referencia para la creación de servicios Web RESTful JAX-RS (JSR 311 y JSR 339), puesto que proporciona una extensa biblioteca para implementar servicios Web RESTful en un contenedor de servlets de Java. Los servlets escanean clases predefinidas para identificar recursos RESTful, analizan la solicitud HTTP entrante, y a continuación seleccionan la clase y el método correctos para responder a dicha solicitud.

Capítulo 6

Generación de adivinanzas a partir de relaciones entre conceptos

Partiendo del proceso detallado en la sección 5.1, se ha desarrollado un generador de adivinanzas que crea acertijos sobre conceptos representados como sustantivos comunes. Estos acertijos se basan en comparaciones entre el concepto destino (por ejemplo, *diamante*) y otras entidades que comparten algunas de sus propiedades (por ejemplo, *cemento*). Los acertijos resultantes se componen como una secuencia de comparaciones utilizando esta plantilla:

“¿Qué es . . . tan *atributo* como *concepto*?”, donde *atributo* es una propiedad del concepto buscado que es la respuesta al acertijo, y *concepto* es una entidad diferente que comparte el valor del atributo con el concepto destino. Por ejemplo:

“¿Qué es . . . tan duro como el cemento y tan transparente como el pelo?”
es un acertijo generado para el concepto diamante por el generador de adivinanzas.

6.1. Generación de adivinanzas

La Tabla 6.1 muestra una serie de ejemplos de los conceptos destino y cómo utilizamos Thesaurus Rex para generar las comparaciones. Los pasos para generar un acertijo siguen el proceso detallado en la sección 5.1.2. A continuación, partiendo de *sun* como primer concepto, se explica más en detalle el proceso.

1. **Categorías del concepto destino.** El primer paso es extraer una lista de todas las categorías mediante una búsqueda en Thesaurus Rex. De esta lista de categorías a las que pertenece el concepto destino, sólo se consideran como candidatas el L % con mayor peso. El valor de L % es configurable (siendo

| <i>Concepto destino</i> | <i>sun</i> | <i>whale</i> | <i>diamond</i> |
|---|--|--|--|
| <i>Categorías</i> | body, object, star... | animal, mammal, predator... | stone, material, gem... |
| <i>Modificadores</i> | stellar, hot , natural, yellow... | large, marine, migratory , aquatic... | precious, hard , valuable, crystalline... |
| <i>Nuevas categorías para el modificador seleccionado</i> | food , beverage... | bird , fish... | material, surface ... |
| <i>Nueva búsqueda</i> | hot food | migratory bird | hard surface |
| <i>Conceptos obtenidos</i> | chili, soup , garlic... | goose , duck, heron... | wood, wall, concrete ... |
| <i>Comparación</i> | as hot as soup? | as migratory as goose? | as hard as concrete? |

Tabla 6.1: Ejemplos de comparaciones obtenidas para ser parte de las adivinanzas. Las palabras en negrita representan las elecciones hechas para cada ejemplo.

en este caso $L = 0.4$). Las categorías con pesos más altos en Thesaurus Rex para el concepto *sun* son *body* y *object*.

2. Selección de modificador

Extracción del modificador. El segundo paso es obtener una lista con los modificadores asociados al concepto destino, a través de una nueva consulta a Thesaurus Rex. De esta lista, los $P\%$ de los atributos con los pesos más altos se consideran como candidatos (en este ejemplo, $P = 0,6$). Para nuestro concepto destino *sun*, algunos de los modificadores extraídos más importantes son: *stellar*, *hot*, *natural* y *yellow*.

Elección de modificador. Uno de los modificadores obtenidos previamente es seleccionado aleatoriamente para evitar un sistema muy repetitivo. Para el ejemplo actual, el sistema elige el modificador *hot*.

3. Selección de categoría

Extracción de categorías. A partir del modificador seleccionado en el paso anterior, se realiza una nueva consulta a Thesaurus Rex, para obtener las categorías que presentan este modificador como una propiedad representativa. Las nuevas categorías seleccionadas podrían ser *food* y *beverage*, que son categorías que presentan la propiedad *hot* en Thesaurus Rex.

Elección de categoría. Se selecciona una de las categorías del paso anterior pero que no esté entre las categorías del concepto destino, que no se encuen-

tre en la lista del primer paso. Para el ejemplo actual, *food* es la categoría seleccionada.

4. **Composición de nueva búsqueda.** Entonces se compone una nueva consulta para Thesaurus Rex utilizando la categoría obtenida en el paso anterior y el modificador seleccionado en el tercer paso. En el ejemplo actual, esta nueva consulta es *hot food*.
5. **Selección de concepto final.** A partir de los conceptos resultantes de la consulta compuesta del paso anterior, se elige un concepto aleatoriamente, en este caso es *soup*.
6. **Plantilla comparación.** Con los datos finales obtenidos durante este proceso, atributo y nuevo concepto, se rellena la plantilla “as *atributo* as *nuevo concepto*?”. El resultado para este ejemplo es “as *hot* as *soup*?”.
7. **Composición adivinanza.** Los pasos 2-6 de este proceso se repiten tantas veces como se desee, lo cual se determina en la configuración del sistema. En cada vuelta se genera una nueva comparación y se añade a la adivinanza final. En este caso, un posible acertijo con tres comparaciones, es el siguiente:

What is ...
... as hot as soup?
... as stellar as a galactic_nucleus?
... as yellow as a mango?

6.2. Evaluación del generador de adivinanzas

Llevamos a cabo una evaluación en esta etapa del desarrollo del sistema para comprobar si las asociaciones de palabras obtenidas proporcionaban información útil para la generación de acertijos, y para evaluar la calidad de los acertijos resultantes. Con este fin, se pidió a un grupo de evaluadores que adivinasen los conceptos iniciales que habían sido utilizados para crear un grupo de adivinanzas, es decir, la solución a dicha adivinanza. A continuación, se estudió la tasa de aciertos obtenida por los evaluadores, a la vez que se analizó cuantas comparaciones eran necesarias para obtener la mayor cantidad de respuestas correctas en los diferentes acertijos. Cuando se crearon los acertijos, fuimos conscientes de la aparición de una serie de problemas relacionados con la ambigüedad y la contradicción de algunas propiedades. Por esta razón se decidió crear dos conjuntos diferentes de adivinanzas para realizar la evaluación, las cuales presentan las comparaciones en diferente orden.

6.2.1. Diseño de la evaluación

Se presentaron diez acertijos a los evaluadores para comprobar si eran capaces de encontrar los conceptos iniciales, es decir, introducidos en el sistema para generar dichos acertijos. Estos acertijos estaban basados en conceptos cotidianos tales como animales, alimentos o vehículos elegidos manualmente. Las adivinanzas se presentaron en cuatro fases, con el fin de saber qué cantidad de comparaciones o pistas necesitaban para resolver cada adivinanza. Para ello, en la primera fase se presentaba una única comparación, en la segunda fase se presentaban dos comparaciones, tres comparaciones en la tercera fase y finalmente, cuatro comparaciones en la cuarta y última fase. No se añadieron más fases puesto que cada una de ellas implica diez acertijos más y pensamos que podría resultar una evaluación larga y tediosa para los sujetos evaluados, corriendo el riesgo de que las respuestas de la últimas fases no fuesen tan fiables.

La evaluación se llevó a cabo utilizando Google Forms y se recogieron algunos datos personales con fines estadísticos (edad, género y habilidad para resolver acertijos). Además, las adivinanzas se tradujeron al castellano para adecuarlas a los usuarios que iban a participar en la evaluación.

Como se explica en el capítulo anterior 5.1, las comparaciones utilizadas en las adivinanzas se eligen aleatoriamente por el sistema. Con ello se quería aportar diversidad a las adivinanzas generadas, de modo que para un mismo concepto no se generase siempre el mismo acertijo. Cuando se generaron los acertijos para la evaluación, nos percatamos de que algunas de las comparaciones no aportaban ningún tipo de información útil a las anteriores comparaciones, o la información añadida era contradictoria o no válida debido en algún caso a la polisemia del concepto. Este es el caso de *coke*, que tanto puede ser una bebida carbonatada con sabor, como el nombre en la calle para la cocaína. Para dicho concepto, nuestro sistema generó las comparaciones “*is as carbonated as ...*” y “*is as hard as ...*”, donde la segunda comparación es obtenida debido a la asociación de palabras “*coke - hard drug*”. Los ejemplos de comparaciones contradictorias, en la mayoría de los casos, están relacionados con atributos con valores imprecisos o difíciles de medir, como el tamaño y la edad. Como ejemplo, el concepto *dog*, Thesaurus Rex lo categoriza con alto valor de peso tanto en *large:animal* como en *small:animal*. En el caso de que nuestro sistema eligiese aleatoriamente ambos atributos para generar la adivinanza, se estarían proporcionando pistas contradictorias.

Para poder realizar una evaluación más detallada y comprobar cómo de valioso era el orden de aparición de las pistas proporcionadas, decidimos crear dos conjuntos diferentes de adivinanzas. Utilizando los mismos diez conceptos, pero con algunas diferencias en las comparaciones proporcionadas, se obtuvo una versión original y una versión revisada de los acertijos. Para el primer conjunto, se selec-

| Conjunto aleatorio | Conjunto revisado |
|--|--|
| <i>... as transparent as hair?</i> | <i>... as hard as concrete?</i> |
| <i>... as pure as gold?</i> | <i>... as transparent as hair?</i> |
| <i>... as expensive as a car?</i> | <i>... as precious as silver?</i> |
| <i>... as simple as a screwdriver?</i> | <i>... as geometrical as a circle?</i> |

Tabla 6.2: Adivinanza A6 de la evaluación a partir del concepto *diamond*

cionaron al azar las cuatro comparaciones resultantes. Para el otro conjunto, se seleccionaron manualmente las cuatro comparaciones más significativas y no contradictorias entre siete comparaciones generadas usando el proceso descrito con el objetivo de evitar las comparaciones no válidas debido a la polisemia o a contradicciones semánticas. En la Tabla 6.2 se muestra el caso concreto de ambos conjuntos para el concepto *diamond*. En el apéndice A se encuentra la lista completa de los acertijos utilizados en la fase final de la evaluación, con el máximo de pistas que se dio.

6.2.2. Resultados de la evaluación

Tanto la versión de evaluación con adivinanzas aleatorias, como aquella con adivinanzas revisadas se presentaron en paralelo a 12 evaluadores diferentes, resultando un total de 24 participantes en el experimento. El orden de aparición de cada acertijo era fijo, de manera que se presentaron las mismas adivinanzas a todos los evaluadores que participaron en cada parte del experimento. Para consultar los resultados de cada acertijo, las Tablas 6.3 y 6.4 muestran los porcentajes de acierto desglosados para cada acertijo en cada fase. Cada fila pertenece a una de las cuatro fases de la evaluación. Los porcentajes de adivinanzas correctamente acertadas en la evaluación por cada fase se presentan en la Tabla 6.5.

| A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | A10 |
|------|-----|------|------|------|------|-----|------|------|-----|
| 17 % | 0 % | 17 % | 8 % | 0 % | 0 % | 0 % | 8 % | 0 % | 0 % |
| 8 % | 0 % | 67 % | 0 % | 33 % | 25 % | 8 % | 17 % | 0 % | 0 % |
| 8 % | 0 % | 58 % | 33 % | 50 % | 92 % | 8 % | 67 % | 0 % | 0 % |
| 8 % | 0 % | 42 % | 17 % | 50 % | 67 % | 0 % | 75 % | 17 % | 0 % |

Tabla 6.3: Porcentaje de acierto para cada adivinanza en el conjunto aleatorio

Las tablas presentadas muestran que para el conjunto revisado se produce un incremento constante en el acierto de las adivinanzas según avanza la fase, es decir, se van proporcionando más pistas. Mientras que para el conjunto generado

| A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | A10 |
|------|------|------|------|------|------|------|------|------|-----|
| 0 % | 25 % | 0 % | 17 % | 0 % | 8 % | 17 % | 0 % | 33 % | 0 % |
| 33 % | 25 % | 75 % | 33 % | 25 % | 58 % | 67 % | 75 % | 33 % | 0 % |
| 42 % | 25 % | 75 % | 42 % | 58 % | 83 % | 75 % | 75 % | 67 % | 8 % |
| 42 % | 33 % | 92 % | 50 % | 58 % | 75 % | 75 % | 75 % | 67 % | 8 % |

Tabla 6.4: Porcentaje de acierto para cada adivinanza en el conjunto revisado

| Fases | Conjunto aleatorio | Conjunto revisado |
|--------|--------------------|-------------------|
| Fase 1 | 5 % | 10 % |
| Fase 2 | 16 % | 43 % |
| Fase 3 | 32 % | 55 % |
| Fase 4 | 28 % | 58 % |

Tabla 6.5: Porcentaje de adivinanzas acertadas

aleatoriamente, el porcentaje de máximo acierto se produce en la tercera fase y luego disminuye ligeramente con la última pista proporcionada.

6.2.3. Discusión

El enfoque que se ha utilizado para generar las adivinanzas asegura que todas ellas tienen solución. En el modo de generación aleatoria, todos los participantes de la evaluación acertaron como mínimo 2 adivinanzas (5 %) y 30 como máximo (32.5 %), con una media de 8 adivinanzas acertadas (20 %) por persona. En la versión de generación revisada, la cantidad mínima de acertijos adivinados fue 6 (15 %) y la cantidad máxima fue 24 (60 %), con una media de 16.5 adivinanzas acertadas (41.3 %) por persona. El objetivo no es conseguir que todos los acertijos sean resueltos, pues indicaría que son demasiado fáciles, no habría ningún tipo de desafío.

Como se muestra en la Tabla 6.5, los resultados de la versión de adivinanzas revisada son significativamente mejores que los de la versión aleatoria. De esta forma, resulta evidente que en algunos casos se necesita una selección de las comparaciones en un orden especial.

Con respecto al número de comparaciones necesarias para acertar la respuesta correcta, se puede observar un hecho curioso en la Tabla 6.4. Con una sola comparación, no hay casi ninguna posibilidad de adivinar el concepto destino. En la mayoría de los casos, la gente responde al azar, pues hay demasiados conceptos que comparten el atributo presentado. Cuando se proporcionan dos comparaciones, los

usuarios son capaces de multiplicar por cuatro el número de respuestas correctas. Cuando se proporcionan tres comparaciones, en el caso de las comparaciones aleatorias, alcanzan su máxima tasa de éxito. En el caso de las comparaciones seleccionadas manualmente, se aciertan el 55 % de las adivinanzas, que es casi el máximo éxito, debido a que la diferencia con la última fase, donde se proporcionan cuatro comparaciones, es casi insignificante.

A un nivel más detallado, las Tablas 6.3 y 6.4 muestran que el acertijo A10 tiene, en el mejor de los casos, una tasa de acierto del 8 %. Esto se debe al hecho de que los atributos seleccionados para describir el concepto destino de esta adivinanza no son lo suficientemente específicos y existen una gran cantidad de propiedades compartidas con otros muchos conceptos. En este ejemplo, el concepto era *avión* y los atributos seleccionados eran: *mecánico*, *rápido*, *móvil* y *complicado*, los cuales son todos atributos comunes.

Sin embargo, el tercer concepto, *sol*, en la cuarta fase del conjunto revisado, tuvo una tasa de acierto del 92 %, como se puede observar en la Tabla 6.4. La razón de este éxito se debe con alta probabilidad a que los atributos seleccionados para este concepto eran mucho más específicos. Por ejemplo, un atributo que no comparten muchos conceptos es *estelar*, lo que combinado con *amarillo*, *caliente* y *central* limita las posibles respuestas a este enigma. Para que se visualice mejor esta idea, la Figura 6.1 muestra las categorías resultantes del atributo más específico utilizado en la adivinanza del concepto *sol*, aquella con mejor tasa de acierto. Mientras tanto, en la Figura 6.2 se pueden observar las categorías obtenidas con uno de los atributos, aquel con menor cantidad de categorías (teóricamente el atributo más específico de los cuatro presentes en el acertijo) del concepto *avión*, el de menor tasa de acierto.

Desde el punto de vista de la tasa de éxito en cada conjunto, en la última fase del conjunto aleatorio (véase la Tabla 6.3), en algunos casos el porcentaje de respuestas correctas disminuye ligeramente. La razón de este suceso, según lo explicado por los participantes en la evaluación, es que a veces las últimas pistas resultaban contradictorias, y los usuarios se encontraban confusos, terminando por cambiar la respuesta en el último momento. Sin embargo, en el conjunto revisado (véase la Tabla 6.4), sólo el acertijo A6 presenta una disminución de la tasa de acierto en la última fase. Esto significa que en este caso al igual que en el ejemplo anterior, la última comparación para algunos de los evaluadores aportaba más confusión que claridad.

Simple categories matching *stellar*:

material, primary, *wind*, **object**, system, property, component, **mass**, motion, structure, medium, **event**, body, phenomenon, activity, quantity, cluster, agglomeration, process, interaction, evolution,

Figura 6.1: Captura del servicio web Thesaurus Rex de las categorías a las que pertenece el atributo *stellar*

Simple categories matching *fast*:

object, food, sound, change, problem, device, block, prey, assistance, activation, vehicle, bus, motorcycle, action, means, picture, reaction, process, component, movement, transport, ship, animal, hound, fire, system, fish, motion, boat, particle, signal, delivery, metal, carnivore, car, weapon, industry, craft, gas, predator, data, fixing, game, conveyance, move, machine, projectile, storage, sport, medium, fare, quarry, thing, phenomenon, bug, structure, equipment, creature, mammal, river, event, memory, bird, meal, critter, entity, shark, loop, insect, activity, sensor, item, storm, cat, subject, vessel,

Figura 6.2: Captura del servicio web Thesaurus Rex de las categorías a las que pertenece el atributo *fast*

6.3. Conclusiones

Se ha presentado un mecanismo automatizado para la generación de adivinanzas usando Thesaurus Rex, un recurso basado en asociaciones de palabras. Siguiendo el procedimiento descrito para generar las pistas, la posterior evaluación señala que las asociaciones de palabras obtenidas por nuestro sistema son útiles para la generación de acertijos. Sin embargo, la evaluación también muestra que sería útil una selección manualmente ordenada de las comparaciones, ya que pueden producirse pistas confusas cuando se trata con conceptos polisémicos o atributos contradictorios, como se ha explicado en la sección 6.2.1. Por lo tanto, sería necesario desarrollar algunos mecanismos para seleccionar sólo los modificadores relacionados con el significado deseado del concepto destino, y considerar de manera especial los atributos con valores confusos.

Los resultados también sugieren que el orden en el que se proporcionan las comparaciones es relevante para resolver el acertijo utilizando menos comparaciones, por lo que puede ser útil analizar el poder de discriminación de cada atributo, de

modo que la complejidad de los acertijos pueda ser controlada. Si esta información estuviese disponible, el sistema podría seleccionar al principio (o al final) los atributos más discriminatorios del concepto de forma automática. La idea subyacente es que cuanto mayor sea el poder de discriminación del atributo, sin llegar a ser un atributo desconocido, más fácil será la adivinanza, puesto que más conceptos se excluyen de las posibles respuestas. Como este razonamiento no se ha cumplido en otros casos, habría que comprobar si para nuestro caso es cierta esta teoría, evaluando con otro conjunto de datos en el que los atributos tengan distintos grados de discriminación. Dependiendo de la dificultad deseada para el acertijo, se podría jugar con el orden de los atributos de acuerdo a su capacidad de discriminación.

Como se ha visto durante la evaluación, en general se necesitan tres o más comparaciones para acertar las adivinanzas. En el futuro, queremos evaluar con cinco o más comparaciones para determinar el número óptimo con el fin de tener acertijos que no son imposibles de adivinar, mientras que al mismo tiempo no son demasiado evidentes. Además, en futuras evaluaciones, nos gustaría incluir acertijos creados por seres humanos, de forma que podamos evaluar si nuestros acertijos son más fáciles o más difíciles en comparación, y si son considerados naturales en comparación con aquellos creados por humanos.

En el futuro, nos gustaría explorar la posibilidad de desarrollar acertijos más creativos, por ejemplo, con rimas y una selección más elaborada de atributos y conceptos. Se podrían utilizar para embellecer el texto a modo de lenguaje figurativo. Un ejemplo de ello sería:

Dentro de una vaina voy y ni espada ni sable soy

Una de nuestras principales preocupaciones a la hora de desarrollar el generador de adivinanzas descrito es la aplicación práctica del sistema resultante. Uno de los objetivos actuales con la generación de acertijos es estudiar la relación entre los conceptos a través de propiedades compartidas para obtener una visión más profunda que nos ayude a generar mejores recursos lingüísticos.

Otra línea de investigación en la que hemos pensado explorar la aplicabilidad de las técnicas descritas es la accesibilidad, y más específicamente, la simplificación de texto y la generación de texto a los usuarios con discapacidad cognitiva. Otros autores ya han tratado sobre el uso de adivinanzas para permitir a los niños con dificultades de comunicación desarrollar sus habilidades lingüísticas (Manurung et al., 2008). Siguiendo esta idea, nuestro objetivo es explorar la forma en que las adivinanzas se pueden incorporar en la vida de las personas con discapacidades de comunicación, con el apoyo, además, del uso de pictogramas.

En base a este trabajo se redactó un artículo para Language Resources and Evaluation Conference 2016. Está aceptado bajo el título *Riddle Generation using Word Associations* y en el Apéndice C se puede encontrar la referencia completa.

Capítulo 7

Generación de figuras retóricas a partir de relaciones entre conceptos

El uso del lenguaje figurativo (véase la sección 4.1) es una herramienta fundamental en la comunicación lingüística. Una de las características más fácilmente identificables del texto generado por ordenador es la tendencia a adherirse a los significados literales. Esto es en parte porque los significados literales son inequívocos y tienen menos riesgo de mala interpretación. Pero también es en parte debido al hecho de que en la actualidad tenemos modelos muy pobres de cómo poder construir el lenguaje figurativo en términos de cálculo.

Disponemos de una cantidad enorme de figuras literarias, entre ellas hay algunas que transforman el texto en un uso incorrecto, que suprimen partes necesarias del contenido, que alteran el orden normal de las oraciones, o que utilizan la repetición para un uso no convencional del lenguaje. Puesto que nosotros buscábamos recursos en los que las asociaciones de palabras jugaran un papel importante, nos hemos centrado en el grupo de figuras retóricas que pueden englobarse en metáforas, y que incluyen símiles y analogías. Este grupo se caracteriza por realizar una comparación entre dos conceptos, a través de un desplazamiento o transferencia de propiedades.

- En la *metáfora* *Booger fue un león en la arena electoral*, se transfieren las cualidades de *león* (la fuente) a Booger (el destino).
- Un *símil* es una especie de metáfora donde se hace la comparación utilizando el nexa “como”. Por ejemplo, *Booger era como un león*.
- En la *analogía* *Booger fue tan valiente como un león*, la cualidad de ser *valiente* (propiedad) se utiliza para vincular *león* (la fuente) a *Booger* (el destino).

7.1. Generación de figuras retóricas

El enfoque propuesto recibe un nombre común como entrada, que es el concepto destino para el que se debe generar la asociación de palabras. Utilizando Thesaurus Rex, el sistema crea una comparación entre el concepto destino y otro concepto que actúa como fuente de la figura retórica, con propiedades similares al concepto destino con el fin de crear una asociación de palabras.

A continuación, se muestran los pasos para generar las asociaciones para las figuras retóricas de uno de los ejemplos de la evaluación basado en un concepto abstracto y cuyos conceptos fuente y destino corresponden a la misma categoría. Para ver el diagrama de flujo que representa este proceso, véase la Figura 5.1. La Tabla 7.1 contiene los datos de dos conceptos abstractos de la evaluación.

| <i>Concepto destino</i> | <i>government</i> | <i>saving</i> |
|---|---|---|
| <i>Conceptos fuente y destino de</i> | misma categoría | distinta categoría |
| <i>Categorías</i> | institution, sector organization... | process, behavior, practise... |
| <i>Modificadores</i> | social , powerful organized, influential... | economic, important, productive , financial... |
| <i>Nuevas categorías para el modificador seleccionado</i> | group, activity organization, event, institution ... | activity , resource, use, behavior, assets... |
| <i>Nueva búsqueda</i> | social institution | productive activity |
| <i>Conceptos obtenidos</i> | church, religion family ... | agriculture, farming fishing... |

Tabla 7.1: Ejemplos de asociaciones obtenidas para formar las figuras retóricas. Las palabras en negrita representan las elecciones hechas para cada ejemplo.

1. **Categorías del concepto destino.** El primer paso es la extracción de todas las categorías del concepto destino mediante una búsqueda en Thesaurus Rex. Sólo se consideran como candidatas el L % con mayor peso. El valor de L % es configurable (en este caso $L = 0.4$). Las categorías con pesos más altos en Thesaurus Rex para *government* son *institution*, *sector* y *organization*.

2. **Selección de modificador**

Extracción del modificador. El segundo paso es obtener los modificadores asociados al concepto destino, a través de una nueva consulta a Thesaurus Rex. De esta lista, los P % de las propiedades con los pesos más altos se

consideran como candidatos (en este ejemplo, $P = 0,8$). Para nuestro concepto destino *government*, algunos de los modificadores extraídos más importantes son: *social*, *powerful*, *organized* y *influential*.

Elección de modificador. Uno de los modificadores obtenidos previamente es seleccionado aleatoriamente para evitar un sistema muy repetitivo. Para el ejemplo actual, el sistema elige el modificador *social*.

3. Selección de categoría

Extracción de categorías. A partir del modificador seleccionado en el paso anterior, se realiza una nueva consulta a Thesaurus Rex, para obtener las categorías que presentan este modificador como una propiedad representativa. Las categorías con más peso para el modificador *social* en Thesaurus Rex son *group*, *activity*, *event* e *institution*.

Elección de categoría. Como en este caso estamos siguiendo el ejemplo en que concepto fuente y destino deben pertenecer a la misma categoría, de las nuevas categorías obtenidas en el paso anterior, se selecciona la última propiedad *institution* que también se presenta en las categorías del primer paso para el concepto *government*.

4. **Composición de nueva búsqueda.** Entonces se compone una nueva consulta para Thesaurus Rex utilizando la categoría obtenida en el paso anterior y el modificador seleccionado en el tercer paso. La consulta en este ejemplo es *social institution*.
5. **Selección de concepto final.** A partir de los conceptos resultantes de la consulta compuesta del paso anterior, se elige un concepto aleatoriamente, en este caso es *family*.
6. **Composición figuras retóricas.** Se generan las tres figuras para esta asociación de palabras:

Government is family
Government is like family
Government is as social as family

7.2. Evaluación del generador de figuras retóricas

El objetivo de esta evaluación ha sido doble. Por un lado, estaba la intención de comprobar la validez de las analogías, símiles y metáforas generadas por nuestro sistema, con el fin de ser capaces de refinar el proceso seguido para generarlos.

Por otro lado, también esperábamos averiguar qué tipo de figura retórica es más evocadora para los evaluadores y cuáles están más cerca de una figura retórica generada por seres humanos.

Este enfoque utiliza la estructura más pura y simple de analogías, símiles y metáforas:

- Analogía: *DESTINO es tan PROPIEDAD como FUENTE.*
- Símil: *DESTINO es como FUENTE.*
- Metáfora: *DESTINO es FUENTE.*

7.2.1. Diseño de la evaluación

El conjunto de evaluación estaba compuesto por 36 analogías, 36 símiles y 36 metáforas. Para crear estos elementos, se utilizaron 36 palabras diferentes como conceptos destino y se crearon una analogía, un símil y una metáfora para cada uno de ellos. Con el fin de evitar la posibilidad de que un evaluador evaluase varias figuras retóricas relacionadas con el mismo concepto destino, el conjunto de datos original se dividió en tres subconjuntos diferentes de 36 figuras retóricas. Cada subconjunto tenía 12 metáforas, 12 símiles y 12 analogías, todos ellos creados a partir de un concepto destino diferente.

Según exponen Jakoff y Johnson en lingüística cognitiva (Lakoff y Johnson, 1980), la mayoría de nuestros conceptos son abstractos como el tiempo, las emociones, la comunicación, la mente, las ideas, las instituciones o las relaciones interpersonales. Enfatizan el hecho de que habitualmente los conceptos abstractos se definen metafóricamente en términos de conceptos más concretos y más estructurados en sus propios términos como son el espacio, el movimiento, la alimentación, los objetos, etc. Defienden que los conceptos abstractos suelen definirse metafóricamente en términos de más de un concepto concreto y, por lo tanto, cada metáfora define sólo ciertos aspectos de un concepto abstracto. En base a este razonamiento, hemos querido evaluar un grupo de figuras retóricas basadas en conceptos abstractos como concepto destino y otro grupo basado en conceptos concretos.

La evaluación se llevó a cabo como una encuesta en línea usando Google Forms, donde cada evaluador recibió un enlace a una de las tres encuestas y se le pidió puntuar cada una de las figuras utilizando la escala Likert. Esta es una escala psicométrica utilizada habitualmente en cuestionarios y es la escala más usada en encuestas de investigación. Para responder una pregunta de un cuestionario elaborado con dicha técnica, se especifica el nivel de acuerdo o desacuerdo con un elemento o pregunta, para conocer el grado de conformidad del encuestado. Las respuestas se pueden presentar en diferentes niveles de medición, permitiendo

escalas de 5, 7 y 9 elementos. Siempre debe haber un elemento neutral para aquellos usuarios que no estén de acuerdo ni en desacuerdo. Esto se debe al hecho de ser una escala que mide actitudes, es decir, que debe aceptar que las personas tienen actitudes favorables, desfavorables o neutras hacia las cosas. Por ello es importante considerar siempre que una escala de actitud debe estar abierta a la posibilidad de aceptar respuestas neutrales. En nuestro caso, se pidió a los evaluadores evaluar cómo de apropiada era o cómo de natural sonaba cada figura retórica, dándoles una puntuación del 1 al 7 (donde 1 simboliza un tropo totalmente inadecuado y 7 representa un tropo que suena completamente natural).

A la hora de interpretar los resultados del cuestionario, cada elemento se puede analizar por separado o, se pueden sumar las respuestas a cada elemento para obtener una puntuación total para un grupo de elementos, puesto que las escalas Likert son escalas sumativas. A su vez, son escalas de tipo ordinal, y al ofrecer valores ordinales, es posible calcular la mediana y la moda, pero no la media ni la desviación típica. La interpretación de la media aritmética en la gestión de categorías tales como “totalmente significativo” o “totalmente sin sentido”, no proporcionaría información útil. Si sumásemos un valor de “totalmente significativo” (5) a dos valores “sin sentido” (2) obtendríamos un promedio de 4, pero eso no es una interpretación muy enriquecedora. De modo que se optó por utilizar la mediana y la moda, ya que cuando se trabaja con escalas Likert son las métricas más interesantes.

Con el fin de tener dos líneas base diferentes en nuestro experimento para medir la calidad de las figuras retóricas generadas por nuestro sistema, hemos utilizado un conjunto de figuras retóricas comúnmente aceptadas, junto con un conjunto de figuras aleatorias manualmente generadas, para compararlas con las generadas por nuestro sistema. Del conjunto de figuras comúnmente aceptadas, las basadas en conceptos abstractos fueron extraídas del libro de Lakoff y Johnson (1980), y las basadas en conceptos concretos se obtuvieron de la web. Además, también por el hecho de explorar diferentes variaciones, las figuras generadas por nuestro sistema se categorizan en dos grupos, en función de si los conceptos generados o fuente pertenecen o no a algunas de las categorías a las que pertenecen los conceptos destino en la aplicación web que consultamos.

En el apéndice B se muestra de manera específica la forma en que se han desarrollado las figuras retóricas de la evaluación.

7.2.2. Resultados de la evaluación

La evaluación fue realizada por 72 evaluadores, de manera que cada uno de los tres subconjuntos de figuras retóricas se evaluó por 24 evaluadores diferentes.

En general, los resultados obtenidos en la evaluación fueron los esperados: las

| Tipo | Moda / Mediana | | |
|---------------------------------------|----------------|----------|-------|
| | Abstracto | Concreto | Total |
| Aleatorio | 1 / 1 | 1 / 1 | 1 / 1 |
| Comúnmente aceptado | 7 / 6 | 7 / 5 | 7 / 5 |
| Generado (diferente categoría) | 1 / 2 | 1 / 2 | 1 / 2 |
| Generado (misma categoría) | 7 / 5 | 1 / 4 | 7 / 5 |
| Generado | 1 / 3 | 1 / 3 | 1 / 3 |

Tabla 7.2: Resultados de metáforas

| Tipo | Moda / Mediana | | |
|---------------------------------------|----------------|----------|-------|
| | Abstracto | Concreto | Total |
| Aleatorio | 1 / 2 | 1 / 1 | 1 / 1 |
| Comúnmente aceptado | 7 / 6 | 5 / 5 | 7 / 5 |
| Generado (diferente categoría) | 1 / 2 | 1 / 3 | 1 / 3 |
| Generado (misma categoría) | 7 / 5 | 6 / 4 | 6 / 5 |
| Generado | 1 / 4 | 1 / 3 | 1 / 4 |

Tabla 7.3: Resultados de símiles

| Tipo | Moda / Mediana | | |
|---------------------------------------|----------------|----------|-------|
| | Abstracto | Concreto | Total |
| Aleatorio | 1 / 1 | 1 / 1 | 1 / 1 |
| Comúnmente aceptado | 7 / 6 | 7 / 6 | 7 / 6 |
| Generado (diferente categoría) | 1 / 3 | 7 / 4 | 2 / 4 |
| Generado (misma categoría) | 5 / 4 | 7 / 4 | 7 / 4 |
| Generado | 1 / 4 | 7 / 4 | 7 / 4 |

Tabla 7.4: Resultados de analogías

| Tipo | Moda / Mediana | | |
|---------------------------------------|----------------|----------|-------|
| | Abstracto | Concreto | Total |
| Aleatorio | 1 / 1 | 1 / 1 | 1 / 1 |
| Comúnmente aceptado | 7 / 6 | 7 / 5 | 7 / 6 |
| Generado (diferente categoría) | 1 / 2 | 1 / 3 | 1 / 3 |
| Generado (misma categoría) | 7 / 5 | 7 / 4 | 7 / 5 |
| Generado | 1 / 4 | 1 / 4 | 1 / 4 |

Tabla 7.5: Resultados generales de la evaluación

figuras retóricas aleatorias resultaron ser las que tenían puntuaciones más bajas, con una mediana de 1, y los tropos con puntuaciones más altas fueron los comúnmente aceptados, con una mediana de 6. Independientemente del tipo de figura retórica y si representaban conceptos concretos o abstractos, la mediana de las figuras generadas por nuestro sistema es 4 (3 para aquellos de diferentes categorías y 5 para los que pertenecen a la misma categoría).

Curiosamente, las modas son las mismas para las figuras retóricas de conceptos abstractos y concretos, independientemente de la forma en que se generaron. La moda tanto de los tropos aleatorios como aquellos generados por nuestro sistema con diferentes categorías es 1, y la moda de los comúnmente aceptados y los generados con la misma categoría también coincide, con un valor de 7.

Si echamos un vistazo más de cerca a los subconjuntos de datos de las metáforas, símiles y analogías, se puede observar que el valor de las medianas de todos los datos generados aleatoriamente en los tres conjuntos de datos es de 1. Los resultados son más satisfactorios para las figuras comúnmente aceptadas, con valores medios entre 5 y 6, lo que demuestra que los evaluadores no se arriesgaron concediendo la máxima puntuación.

Cuando continuamos analizando los subconjuntos, podemos ver que los resultados obtenidos para los tropos que pertenecen a diferentes categorías (con variaciones de la mediana entre 2 y 4) son menos prometedores que los obtenidos para los tropos con la misma categoría (con variaciones entre 4 y 5). En el caso de las analogías, la mediana es la misma para las generadas con la misma categoría o con diferentes categorías, con un valor de 4. La diferencia entre las medianas de las figuras retóricas aleatorias y las comúnmente aceptadas fluctúa entre 4 y 5.

Los gráficos muestran los resultados comparativos de las diferentes formas de generar las figuras retóricas: utilizando conceptos concretos y abstractos, así como los resultados combinados. El primer gráfico (véase la Figura 7.1), se corresponde con la asociación de palabras utilizando conceptos abstractos y se puede observar que las figuras retóricas aleatorias obtienen 1 en todos los resultados, excepto en el caso de la mediana de los símiles, que es 2. La moda de los tropos comúnmente aceptado es 6 y la mediana es 7. La moda de las figuras retóricas generadas con diferentes categorías es siempre 1, mientras que los resultados de la mediana oscilan entre 2 y 3. La moda para los tropos generados con la misma categoría está entre 5 y 7, y la mediana entre 4 y 5.

Los resultados de los conceptos concretos se muestran en la Figura 7.2. De manera similar a los conceptos abstractos, tanto los resultados de la moda como los de la mediana de las figuras retóricas aleatorias son 1. La moda y la mediana de los tropos comúnmente aceptados oscilan entre 5 y 7. La moda de las figuras retóricas generadas con diferentes categorías es 1, a excepción de las analogías, que

es 7. La mediana en este caso se encuentra entre 2 y 4. La mediana de las figuras retóricas con la misma categoría es siempre 4, mientras que la moda es 1 para las metáforas, 6 para los símiles y 7 para las analogías.

En la Figura 7.3 se pueden observar los resultados totales para todas las figuras retóricas. Claramente, aquellas generadas aleatoriamente obtienen los peores resultados con un 1 tanto en la moda como en la mediana. La moda de los tropos comúnmente aceptados es 7, mientras que la mediana es 5 para las metáforas y los símiles, y 6 para las analogías, siendo las comúnmente aceptadas el conjunto de figuras mejor valoradas. La moda de las figuras retóricas generadas con diferentes categorías es 1 y 2, y la mediana es 2 para las metáforas, 3 para los símiles y 4 para las analogías. Tanto la moda como la mediana de las figuras generadas en base a conceptos de la misma categoría consiguen mejores resultados que aquellos basados en conceptos de diferente categoría. La moda de los tropos generados con la misma categoría es 7 para las metáforas y las analogías, mientras que para los símiles es 6. La mediana es 5 para las metáforas y los símiles, y 4 para las analogías.

Podemos concluir que, si bien el proceso que hemos utilizado para generar las figuras retóricas funciona bastante bien cuando se utilizan conceptos de la misma categoría, de acuerdo con las opiniones de los evaluadores, algo diferente sucede en el caso de la utilización de conceptos que pertenecen a diferentes categorías, lo cual, en general, produce peores resultados. Este hecho apunta a la necesidad de utilizar propiedades o relaciones adicionales con el fin de obtener conceptos que posteriormente pueden dar lugar a figuras retóricas más significativas.

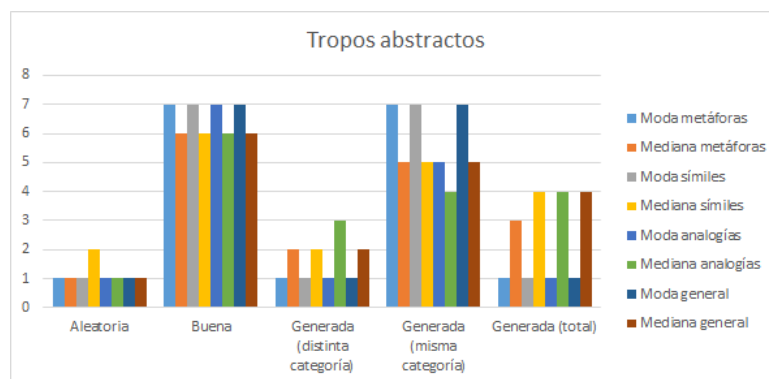


Figura 7.1: Gráfico de barras con los resultados de los conceptos abstractos de las figuras retóricas

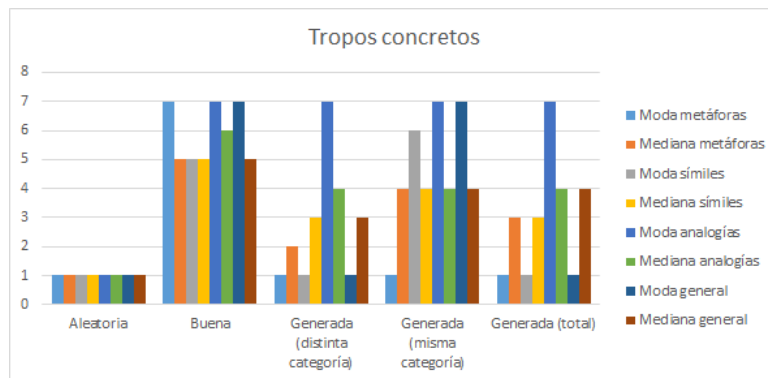


Figura 7.2: Gráfico de barras con los resultados de los conceptos concretos de las figuras retóricas

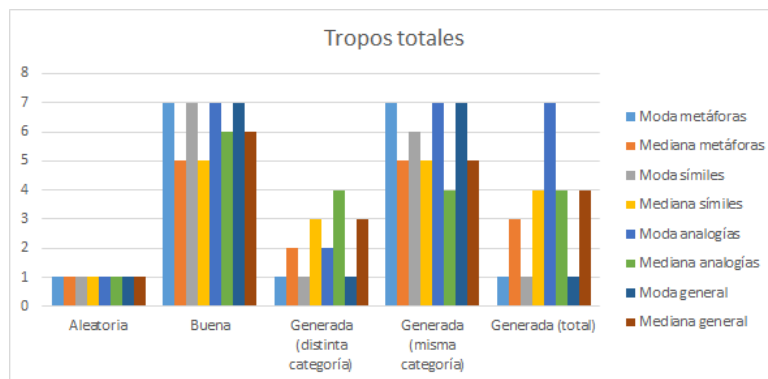


Figura 7.3: Gráfico de barras con los resultados totales de las figuras retóricas

7.2.3. Discusión

Los resultados de la evaluación de las metáforas se muestran en la Tabla 7.2. Como se puede ver, en todos los casos las metáforas generadas aleatoriamente se han valorado como sin sentido por los evaluadores. Por el contrario, las metáforas comúnmente aceptadas consiguen los resultados más altos, con una ligera preferencia por las metáforas creadas utilizando conceptos abstractos sobre las que se basan en el uso de conceptos concretos. Las metáforas generadas automáticamente utilizando conceptos de diferentes categorías también resultan mal clasificadas, lo que señala que compartir una sola propiedad no es suficiente para generar una buena metáfora. Para las metáforas generadas utilizando conceptos que pertenecen a la misma categoría, la diferencia que existe entre las modas de las metáforas que utilizan conceptos concretos y abstractos es notable. Esto sugiere que las metáforas de conceptos abstractos son más evocadoras y ofrecen una gama más amplia de interpretaciones que las de conceptos concretos. Por último, la mediana global de

las metáforas también sugiere que deben tomarse en consideración más aspectos para aumentar la calidad percibida de estas figuras retóricas.

La Tabla 7.3 muestra los resultados para los símiles evaluados. Las valoraciones en este caso son bastante similares a los resultados obtenidos para las metáforas. Aunque en el caso de los símiles generados automáticamente hay una valoración más alta para símiles creados utilizando conceptos concretos que pertenecen a la misma categoría. Esto puede ser debido al hecho de que, en estos símiles, no hay una coincidencia directa entre los conceptos fuente y destino, de manera que los evaluadores son más propensos a aceptar la similitud entre los dos que en el caso de las metáforas.

Los resultados de la evaluación de las analogías se pueden ver en la Tabla 7.4. Las valoraciones en este caso son ligeramente superiores que en las dos figuras retóricas anteriores, probablemente debido al hecho de que el aspecto en el que los dos conceptos son considerados similares se indica explícitamente. Este mismo aspecto puede ser la causa de la menor puntuación obtenida por las analogías generadas automáticamente utilizando conceptos abstractos que pertenecen a la misma categoría. En este caso, la similitud percibida por los evaluadores puede enfocarse en una característica diferente de la elegida por el sistema, lo que causa que la valoración sea inferior que la concedida a las figuras anteriores. Por el contrario, las analogías generadas utilizando conceptos concretos que pertenecen a diferentes categorías están mucho mejor valoradas que en los tropos anteriores. En este caso, la razón parece ser el hecho de que la propiedad utilizada por el sistema para comparar ambos conceptos se ha hecho explícita, de manera que los evaluadores puedan ver la razón por la que el sistema considera los dos conceptos relacionados entre sí y estén más inclinados a aceptarla como válida.

Por último, los resultados globales de la evaluación se pueden ver en la Tabla 7.5. A pesar de que no se diferencian mucho de los resultados obtenidos por los diferentes tropos independientemente, los valores de las modas están claramente desplazados hacia los límites de la escala. Este efecto sugiere que los evaluadores humanos tienden a aceptar o no aceptar una figura retórica como válida, pero las posiciones intermedias son menos comunes. En cuanto al valor de las medianas, los resultados agregados confirman la percepción de que, en términos de figuras retóricas generadas de forma automática, las que utilizan los conceptos abstractos que pertenecen a la misma categoría son apreciadas ligeramente mejor que el resto.

7.3. Conclusiones

Se ha realizado una evaluación de la calidad de las figuras retóricas y se ha demostrado que es posible obtener resultados consistentes. Una de las conclusiones

más claras es que los tropos generados por nuestro sistema de conceptos de la misma categoría tienen mayor calidad que los tropos basados en conceptos de diferentes categorías.

En vista de los resultados, uno de los caminos que tenemos que seguir está dirigido a encontrar maneras de generar buenas figuras retóricas de conceptos de diferentes categorías, porque en la vida cotidiana algunas de las mejores figuras retóricas se construyen a partir de este tipo de términos, tales como *time is money*. Las categorías con mayores pesos obtenidos para el concepto *time* en Thesaurus Rex son *information*, *quantity* y *attribute*, mientras que para el concepto *money* son *thing*, *property*, *value* y *assets*. Como hemos visto en la evaluación, este tropo obtiene una buena calificación y necesitamos más información acerca de este tipo de figuras retóricas.

Con el fin de generar lenguaje figurativo apropiado dependiendo del contenido de un texto dado, sería interesante encontrar conjuntos de palabras agrupadas por temas. Por otra parte, con el fin de adecuar el lenguaje figurado a los objetivos o necesidades del lector, sería útil contar con conjuntos de conceptos agrupados por la complejidad semántica.

A veces, las limitaciones encontradas surgen de la web en sí misma. Esto se debe a que generalmente la información disponible en la web tiende a ser más literal que evocadora. Para el ejemplo anterior, los atributos con más peso obtenidos de Thesaurus Rex cuando se busca el concepto *time* son *physical*, *basic*, *measurable*, *relevant* y *abstract*. Eso sugiere que puede ser más apropiado para nosotros encontrar o generar una fuente de conocimiento específico que proporcione propiedades figurativas.

La moda más alta de las figuras retóricas generadas por nuestro sistema se obtiene de las analogías. En el caso de la mediana para el total de resultados, la tendencia es menos clara. Mientras que para figuras de la misma categoría se producen buenos resultados en todas las figuras, tan sólo en las analogías de conceptos concretos se consiguen buenas valoraciones para figuras de diferentes categorías.

En el futuro, nos gustaría seguir haciendo evaluaciones para encontrar patrones o similitudes entre las figuras retóricas mejor valoradas, y queremos probar esto con grandes conjuntos de datos. De esta manera, los resultados de las evaluaciones podrían servir para mejorar la calidad de los recursos generados por nuestro sistema.

Hemos utilizado los términos “concreto” y “abstracto” para categorizar los conceptos de entrada. Sería interesante comprobar si supone una diferencia utilizar una palabra concreta para describir un concepto abstracto (por ejemplo, “el tiempo es oro”) y viceversa.

Como trabajo futuro también nos gustaría comprobar el grado de similitud

entre el concepto fuente y el concepto destino. Si los conceptos son demasiado similares, la figura retórica resultante sería correcta, pero no muy práctica.

Con respecto a la cantidad de información proporcionada en la figura retórica, no hay diferencias significativas entre las que proporcionan más o menos información, porque se obtienen resultados similares para las metáforas (donde se indican sólo el concepto original y el nuevo concepto) y para las analogías (donde también se muestra el atributo compartido).

Los resultados obtenidos indican que se deben realizar más intentos para mejorar nuestro sistema y generar figuras retóricas de mejor calidad, y así progresivamente tratar de alcanzar los resultados de las figuras retóricas generadas por personas. En el futuro, una característica útil que puede mejorar nuestro sistema es relacionar el concepto original con conceptos que tienen más de una propiedad en común. A partir de ahora, otro camino hacia el que debemos investigar es la generación de figuras retóricas con conceptos que se relacionan a través de dos o más atributos. En el ejemplo *A ballerina is a swan*, ambos conceptos comparten propiedades como *pretty*, *graceful* y *stylized*.

En base a este trabajo se redactó un artículo para International Conference on Computational Creativity 2016. Está aceptado bajo el título *Exploring the Role of Word Associations in the Construction of Rhetorical Figures* y en el Apéndice D se puede encontrar la referencia completa.

Capítulo 8

Conclusiones y trabajo futuro

8.1. Conclusiones

Después de estudiar la forma de utilizar las asociaciones de palabras para generar recursos lingüísticos, se ha comprobado que no es una tarea fácil, puesto que no existen herramientas que refinen, estructuren y relacionen el lenguaje natural con tanta sofisticación como lo hacemos los humanos. Las herramientas relacionadas con la Web Semántica juegan un importante papel para la calidad de nuestro sistema, ya que no hemos implementado ningún modelado de dominios. De forma que, a la hora de generar estas asociaciones, nos hemos encontrado con algunos obstáculos como la polisemia, los adjetivos graduales o los comparativos.

A partir de algunas herramientas disponibles en línea, inicialmente fue Thesaurus Rex, se ha desarrollado un proceso de asociación de palabras, que partiendo de un concepto destino, crea una relación semántica a través de un atributo con un nuevo concepto, el concepto fuente de la asociación, pues el atributo que los relaciona es más representativo en este nuevo concepto. El objetivo de este proyecto era crear recursos lingüísticos partiendo de asociaciones de conceptos, por ese motivo probamos a crear adivinanzas y figuras retóricas, y posteriormente evaluarlas.

Los resultados de la evaluación de adivinanzas señalaron que las asociaciones de palabras obtenidas por nuestro sistema son útiles para la generación de acertijos, pero que sería útil una selección ordenada de las comparaciones, debido a que pueden producirse pistas confusas por los obstáculos que se han comentado, como conceptos polisémicos o atributos contradictorios. También se comprobó que por regla general se necesitan tres o más comparaciones para acertar las adivinanzas. Una buena idea sería evaluar con cinco o más comparaciones para determinar el número óptimo para que los acertijos no sean ni demasiado evidentes, ni imposibles de adivinar, a la vez que se incluyen acertijos creados por seres humanos para establecer una línea base y comparar el grado de dificultad y de naturalidad.

Por otro lado, la evaluación de la calidad de las figuras retóricas demostró que los tropos generados por nuestro sistema de conceptos de la misma categoría tienen mayor calidad que los tropos basados en conceptos de diferentes categorías. Esto significa que fueron mejor aceptadas las figuras cuyos dos conceptos, tanto el concepto fuente, que presta sus atributos, como el concepto destino que los recibe, corresponden a la misma categoría de conceptos. Por consiguiente, debemos dirigir nuestro esfuerzo a generar buenas figuras retóricas de términos de diferentes categorías, porque algunas de las mejores figuras retóricas conocidas y valoradas en nuestra rutina diaria, se construyen a partir de este tipo de conceptos, como se detalla en la sección 7.3 con el ejemplo de la evaluación *time is money*. También hemos sacado la conclusión de que algunas limitaciones se deben a la misma web, puesto que los datos que nos encontramos son en general más literales que evocadores. Lo cual nos sugiere que quizás deberíamos buscar una fuente de conocimiento que proporcione un tipo de propiedades más figurativas.

Se han conseguido superar tanto los objetivos científicos (véase la sección 3.1) como los objetivos tecnológicos (véase la sección 3.2) que se propusieron inicialmente para este trabajo:

- Se ha comprobado la situación actual de las herramientas desarrolladas en el campo de la creatividad computacional y la Generación de Lenguaje Natural como se detalla en el capítulo 4. En base a ese estudio, hemos conocido el panorama actual y la evolución que han tenido estos campos. Pudiendo así, situarnos y saber mejor cómo enfocar nuestro trabajo, refinando así nuestros objetivos.
- También se han explorado específicamente recursos lingüísticos en línea para extraer información que enriquezca nuestro sistema, descritos en la sección 4.2. Primero de todo, y en mayor medida, se ha experimentado con Thesaurus Rex, donde se realizan las principales consultas para generar las asociaciones de conceptos. Posteriormente, se añadieron las herramientas ConceptNet y WordNet.
- Posteriormente a la realización de las evaluaciones de nuestro generador de recursos, hemos discutido los resultados y obtenido conclusiones acerca de la utilidad de las asociaciones de conceptos para dicho propósito. En las secciones 6.3 y 7.3, y también a lo largo de este capítulo se describen en detalle tanto las conclusiones, como las posibilidades de mejora del generador de texto.
- Hemos implementado un generador de recursos basado en un proceso de asociación de palabras agrupando variedad de tecnologías y la arquitectura

del sistema ha funcionado como esperábamos, permitiendo que la consulta al generador sea personalizable y flexible (véase las secciones 5.1, 4.3 y 4.4). Empezamos por integrar las funcionalidades necesarias para interpretar las salidas de los servicios que consultamos para posteriormente encontrar las asociaciones de palabras. Hemos intentado integrarlas de la mejor manera para por último generar texto en lenguaje natural.

- Se realizaron dos evaluaciones con personas a través de los formularios de Google Form, con todo el proceso de corrección y discusión que conllevan, descritas en las secciones 6.2 y 7.2. La primera de ellas se realizó sobre un conjunto de 10 adivinanzas a lo largo de 4 fases. En cada fase sucesiva se proporcionaban más comparaciones o pistas para conocer el número necesario para completar cada adivinanza. La segunda evaluación utilizaba un conjunto de 36 metáforas, 36 símiles y 36 analogías, creados a partir de 36 conceptos. Se evaluaron junto con un grupo de 6 figuras retóricas comúnmente aceptadas y otras 6 figuras aleatorias para establecer unas líneas base en la evaluación. Las conclusiones de las evaluaciones nos han servido para realizar algunas pequeñas mejoras en el generador.
- El servicio web en línea (mencionado en la sección 5.2) está disponible en el servidor Hefesto para el que lo desee utilizar y se ha conseguido que las consultas acepten diferentes parámetros para personalizar al máximo la búsqueda. Nuestro servicio también cuenta con una página de ayuda para saber cómo manejarlo.

8.2. Trabajo futuro

A pesar de que se han conseguido los objetivos propuestos, también es posible plantearse nuevas características. Estas características pueden incluir desde nuevas funcionalidades hasta mayor personalización del sistema:

- Se podrían ir añadiendo más herramientas semánticas para comparar la fuerza de las relaciones entre conceptos y qué atributos son más frecuentes para cada concepto. Por el momento, nuestro sistema cuenta con las herramientas Thesaurus Rex, WordNet y ConceptNet, pero podríamos incorporar por ejemplo consultas a corpus como los que se pueden encontrar en Sketch Engine ¹ o en Wacky corpora ².

¹<https://the.sketchengine.co.uk/open/>

²https://corpora.dipintra.it/public/run.cgi/first_form

- Sería muy positivo establecer un sistema de valoraciones, tanto para las adivinanzas como para las figuras retóricas, de forma que recibiríamos información sobre la calidad de los recursos generados y serviría para seguir mejorando el sistema. Además, en base a las evaluaciones realizadas se han hecho algunos cambios en el sistema que no se han podido evaluar posteriormente.

Por ejemplo, en la generación de adivinanzas, ahora existe la posibilidad de que el orden de aparición de las pistas o comparaciones dependa del poder de discriminación de cada atributo, es decir que se proporcionen primero los atributos más discriminatorios. Pensamos que si se utilizaban propiedades más específicas, la cantidad de posibles conceptos resultantes disminuiría, por ende, cuánto más descriptivas fueran las comparaciones suministradas, antes serían resueltos los acertijos. Por ejemplo, sería más esclarecedor si el sistema comienza indicando *gravitational, hot, central, yellow, natural* en este orden que en orden inverso.

Por otro lado, en la generación de figuras retóricas serviría para encontrar patrones o similitudes entre aquellas mejor valoradas. Y si realizásemos evaluaciones con grandes conjuntos de datos, los resultados podrían servir en gran medida para mejorar la calidad de los recursos generados por nuestro sistema y progresivamente tratar de alcanzar los resultados de las figuras retóricas generadas por personas.

- Un punto interesante a desarrollar para la generación de adivinanzas sería utilizar relaciones no sólo de propiedad (relación *HasProperty* en Concept-Net), sino también de otros tipos como por ejemplo *MadeOf*, *IsUsedFor* o *IsPartOf*. Con este punto extra, se podría jugar con la creatividad e inventar otro tipo de acertijos más embellecidos, como por ejemplo *¿Qué tiene cuello pero no cabeza?*. Para llegar a crear este acertijo deberíamos jugar con las relaciones *PartOf* de botella (el concepto proporcionado en la petición) e *IsPartOf* de cuello.
- Una línea de investigación que se puede explorar para aplicar nuestro sistema de acertijos es la accesibilidad, concretamente, la simplificación de texto y la generación de texto para los usuarios con discapacidad cognitiva. Como ya se explicó en la sección 7.3, otros autores han investigado su uso con niños con dificultades de comunicación. El objetivo sería explorar el modo de incluir los acertijos en la vida de estas personas con discapacidades de comunicación.
- Una característica útil que también podría mejorar nuestro sistema sería relacionar el concepto introducido con conceptos que tienen más de un atributo en común. Este camino ya se está investigando, generar figuras retóricas con

conceptos que se relacionan a través de dos o más atributos. En el ejemplo *A ballerina is a swan*, ambos conceptos comparten propiedades como *pretty*, *graceful* y *stylized*.

- En el futuro, otra de las funcionalidades que se puede implementar es la polaridad a la hora de crear las asociaciones de conceptos. Es decir, el atributo que relaciona los conceptos, podría hacerlo de manera positiva (por ejemplo, *brillante*), de forma negativa (por ejemplo, *abrasador*) o neutra (como *amarillo*). Esto se podría conseguir a partir del sistema Metaphor Magnet, como se puede ver en la Figura 8.1, en la que se muestran las propiedades positivas y negativas correspondientes al concepto *sun*.

Source Metaphors: +*sun*

shining:diamond, soothing:tenderness, shining:testament, shining:moonlight,
 shining:star, shining:smile, shining:saint, shining:flame,
 shining:angel, shining:victory, beautiful:child, shining:silver_flame,
 shining:pearl, sparkling:watercolor, shining:crown,
 shining:emerald, shining:sunlight, soothing:sauna,
 glowing:sunset, loving:lover, sparkling:falls, soothing:drug, loving:kiss,

Source Metaphors: -*sun*

burning:sunlight, burning:fire, burning:cauldron,
 threatening:giant, burning:flame, damaging:flood,
 threatening:hurricane, burning:desert, screaming:hog, screaming:baby,
 burning:star, damaging:pesticide, smoldering:dream, killing:Dracula,
 burning:fireball, pounding:monsoon, acidic:orange, pounding:waterfall,
 burning:volcano, burning:cigarette_butt, threatening:obstacle,

Figura 8.1: Resultados positivos y negativos del concepto *sun* en Metaphor Magnet

Chapter 9

Conclusions and future work

9.1. Conclusions

After studying how to use word associations to generate linguistic resources, it has been proven that it is not an easy task, since there are no tools that refine, structure and relate natural language with as much sophistication as we humans do. The tools related to the Semantic Web play an important role for the quality of our system, because we have not implemented any domain modelling. So, at the moment of generating these associations, we have met some obstacles such as polysemy, gradual or comparative adjectives.

From some available online tools, initially only Thesaurus Rex, a process of word association has been developed, starting from a target concept, creates a semantic relationship through an attribute with a new concept, the source concept of the association, because the attribute that relates them is more representative in this new concept. The objective of this project was to create linguistic resources based on concept associations, for that reason we try to create riddles and rhetorical figures, and then evaluate them.

The results from the riddle assessment indicated that the word associations obtained by our system are useful for riddle generation, but that an orderly selection of the comparisons would be useful, because there may be confusing hints due to obstacles commented as polysemic concepts or contradictory attributes. It was also found that, as a rule, three or more comparisons are needed to guess the riddles. A good idea would be to evaluate with five or more comparisons to determine the optimum number so that puzzles are neither too obvious nor impossible to guess, while including puzzles created by humans to establish a baseline and compare grade of difficulty and naturalness.

On the other hand, the evaluation of the rhetorical figures quality showed that

the tropes generated by our system of concepts of the same category have higher quality than the tropes based on concepts of different categories. This means that figures whose two concepts, both the source, which lends its attributes, and the target concept that receives them, correspond to the same category of concepts were better accepted. Therefore, we must direct our effort to generate good rhetorical figures of concepts of different categories, because some of the best rhetorical figures known and valued in our daily routine, are constructed from this type of concepts, as detailed in section 7.3 with the example of the evaluation *time is money*. The assessment also led us to the conclusion that some limitations are due to the web itself, since the data that we find are in general more literal than evocative. Which suggests that perhaps we should look for a knowledge source that provides a kind of more figurative properties.

Both the scientific objectives (see section 3.1) and the technological objectives (see section 3.2), which were originally proposed for this work, have been overcome:

- The current situation of the tools developed in the field of computational creativity and the Natural Language Generation has been checked as detailed in the chapter 4. Based on this study, we have seen the current panorama and the evolution of these fields. This has allowed us to situate ourselves and to know better how to focus our work, thus refining our objectives.
- We have also specifically explored online linguistic resources to extract information that enriches our system, described in section 4.2. First of all, and to a greater extent, we have been experimented with Thesaurus Rex, where the main queries are realized to generate the concept associations. Later, the tools ConceptNet and WordNet were added.
- Subsequent to the realization of the evaluations of our resource generator, we have discussed the results and obtained conclusions about the usefulness of the concept associations for this purpose. In sections 6.3 and 7.3, and throughout this chapter both the conclusions and the possibilities for improving the text generator are described in detail.
- We have implemented a resource generator based on a word association process grouping a variety of technologies and the system architecture has worked as expected, allowing the query to the generator to be customizable and flexible (see sections 5.1, 4.3 and 4.4). The first thing was to integrate the functionalities necessary to interpret the outputs of the services that we consulted and later to find the word associations. We have tried to integrate them in the best way to finally generate text in natural language.

- Two evaluations were conducted with people through the Google Form forms, with all the process of correction and discussion they involve, described in sections 6.2 and 7.2. The first one was carried out on a set of 10 riddles over 4 phases. In each successive phase, more comparisons or clues were provided to know the number needed to complete each riddle. The second evaluation used a set of 36 metaphors, 36 similes and 36 analogies, created from 36 concepts. They were evaluated together with a group of 6 commonly accepted rhetorical figures and 6 other random figures to establish baselines in the evaluation. The conclusions of the evaluations have served us to make some small improvements in the generator.
- The online web service (mentioned in section 5.2) is available on Hefesto server for those who want to use it and queries accept different parameters to personalize the search to the maximum. Our service also has a help page to know how to handle it.

9.2. Future work

Although the proposed objectives have been achieved, it is also possible to consider new features. These features can range from new functionalities to greater customization of the system:

- More semantic tools could be added to compare the strength of relationships between concepts and which attributes are most prevalent for each concept. At the moment, our system includes the tools Thesaurus Rex, WordNet and ConceptNet, but we could incorporate for example queries to corpus like those that can be found in Sketch Engine ¹ or in Wacky corpora ².
- It would be very positive to establish an evaluation system, both for riddles and rhetorical figures, so that we would receive information on the quality of the resources generated and would serve to further improve the system. In addition, based on the evaluations carried out, some changes have been made to the system that could not be evaluated later.

For example, in the riddle generation, it is possible for the order of occurrence of the hints or comparisons to depend on the discriminating power of each attribute, i.e. the most discriminating attributes are first provided. We thought that if more specific properties were used, the number of possible concepts that would result would decrease, so how much more descriptive

¹<https://the.sketchengine.co.uk/open/>

²https://corpora.dipintra.it/public/run.cgi/first_form

were the comparisons provided, before the puzzles would be solved. For example, it would be more enlightening if the system begins by indicating gravitational, hot, central, yellow, natural in this order rather than in reverse order.

On the other hand, in the generation of rhetorical figures would serve to find patterns or similarities among those best valued. And if we conducted assessments with larger datasets, the results could serve to improve the quality of the resources generated by our system and progressively try to reach the results of rhetorical figures generated by people.

- An interesting point to develop for riddle generation would be to use relations not only of property (HasProperty relation in ConceptNet), but also of other types such as MadeOf, IsUsedFor or IsPartOf. With this extra point, it could be possible to play with creativity and invent other types of more embellished puzzles, such as *What has neck but not head?*. To create this puzzle, we should play with bottle *PartOf* relationships (the concept provided in the request) and *IsPartOf* of neck.
- One line of research that can be explored to apply our riddles system is the accessibility, specifically, text simplification and text generation to users with cognitive disability. As already explained in section 7.3, other authors have investigated its use with children with communication difficulties. The goal would be to explore how to include riddles in the lives of these people with communication disabilities.
- A useful feature that could also improve our system would be to relate the concept introduced with concepts that have more than one attribute in common. This path is already being investigated, generating rhetorical figures with concepts that are related through two or more attributes. In the example *A ballerina is a swan*, both concepts share properties like pretty, graceful and stylized.
- In the future, another of the functionalities that can be implemented to create concept associations is the polarity. That is, the attribute that relates the concepts, could do so in a positive (e.g., bright), negative (for example, burning) or neutral form (yellow). This could be achieved from the Metaphor Magnet system, as can be seen in Figure 8.1, which shows the positive and negative properties corresponding to the concept sun.

Bibliografía

- BINSTED, K. y RITCHIE, G. Computational rules for generating punning riddles. *Humor - International Journal of Humor Research*, vol. 10(1), páginas 25–76, 1997.
- BLACK, M. Xii.metaphor. *Proceedings of the Aristotelian Society*, vol. 55(1), páginas 273–294, 1955.
- COHEN, H. The further exploits of aaron, painter. *Stanford Humanities Review*, vol. 4, páginas 141–158, 1995.
- COLTON, S. Automated puzzle generation. En *Proceedings of the AISB'02 (Symposium on AI and Creativity in the Arts and Science)*. 2002a.
- COLTON, S. *Automated Theory Formation in Pure Mathematics*. Springer series on distinguished dissertations, 2002b.
- COLTON, S. The painting fool: Stories from building an automated painter. *Computers and creativity*, páginas 3–38, 2012.
- COPE, D. Experiments in music intelligence. En *International Computer Music Conference Proceedings*, páginas 174–181. 1987.
- CUNNINGHAM, H., MAYNARD, D. y TABLAN, V. Jape: a java annotation patterns engine (second edition). Research Memorandum CS-00-10, Department of Computer Science, University of Sheffield, 2000.
- DE PALMA, P. y WEINER, E. Riddles: accessibility and knowledge representation. En *Proceedings of Coling'92*, páginas 1121–1125. 1992.
- DE PALMA, P. y WEINER, E. Some pragmatic features of lexical ambiguity and simple riddles. En *Language & communication*, páginas 183–193. 1993.
- DIAZ-JEREZ, G. Composing with melomics: Delving into the computational world for musical inspiration. *Leonardo Music Journal*, páginas 13–14, 2011.

- GERVÁS, P. An expert system for the composition of formal spanish poetry. *Journal of Knowledge-Based Systems*, vol. 14(3–4), páginas 181–188, 2001.
- GUERRERO, I., VERHOEVEN, B., BARBIERI, F., MARTINS, P. y PÉREZ Y PÉREZ, R. Theriddlerbot: A next step on the ladder towards creative twitter bots. En *Proceedings of the Sixth International Conference on Computational Creativity*, páginas 315–322. 2015.
- HARMON, S. Figure8: A novel system for generating and evaluating figurative language. En *Proceedings of the Sixth International Conference on Computational Creativity*, páginas 71–77. 2015.
- HERVÁS, R., COSTA, R., COSTA, H., GERVÁS, P. y PEREIRA, F. Enrichment of automatically generated texts using metaphor. En *6th Mexican International Conference on Artificial Intelligence (MICA-07)*, vol. 4827, páginas 944–954. Springer Verlag, LNAI Series, Springer Verlag, LNAI Series, 2007.
- HERVÁS, R., PEREIRA, F., GERVÁS, P. y CARDOSO, A. Cross-domain analogy in automated text generation. En *3rd Joint Workshop on Computational Creativity, dentro de la 17th European Conference on Artificial Intelligence*, páginas 43–48. 2006a.
- HERVÁS, R., PEREIRA, F., GERVÁS, P. y CARDOSO, A. A text generation system that uses simple rhetorical figures. *Procesamiento de Lenguaje Natural*, vol. 37, páginas 199–206, 2006b.
- JORDÁN, A. Lenguas y tecnologías de la información. En *La lengua española y las nuevas tecnologías, Inteligencia Artificial y Lengua Española, dentro del Congreso de la Lengua Española en Sevilla 1992*. 1992.
- LAKOFF, G. The contemporary theory of metaphor. *Metaphor and Thought*, páginas 202–251, 1993.
- LAKOFF, G. y JOHNSON, M. The metaphorical structure of the human conceptual system. *Cognitive Science*, vol. 4(2), páginas 195–208, 1980.
- MACHADO, P. y CARDOSO, A. Nevar the assessment of an evolutionary art tool. En *Proceedings of the AISB00 Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science*, vol. 456. 2000.
- MANURUNG, R., RITCHIE, G., PAIN, H., WALLER, A., O'MARA, D. y BLACK, R. The construction of a pun generator for language skills development. *Applied Artificial Intelligence*, vol. 22(9), páginas 841–869, 2008.

- MCCUNE, W. A davis-putnam program and its application to finite first-order model search. Technical Memorandum 194, Mathematics and Computer Science Division, Argonne National Laboratory, 1994a.
- MCCUNE, W. Otter 3.0 reference manual and guide. Informe Técnico ANL-94/6, Mathematics and Computer Science Division, Argonne National Laboratory, 1994b.
- MEEHAN, J. Tale-spin, an interactive program that writes stories. En *Proceedings of International Joint Conference on Artificial Intelligence*, páginas 91–98. 1977.
- PEÑA, A. Qué es un servicio restful. 2016. [Online; accessed 6-September-2017].
- PÉREZ Y PÉREZ, R. y SHARPLES, M. Mexica: A computer model of a cognitive account of creative writing. *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 13, páginas 119–139, 2001.
- PINTÉR, B., VOROS, G., SZABO, Z. y LORINCZ, A. Automated word puzzle generation using topic models and semantic relatedness measures. En *Joint Conference on Mathematics and Computer Science (MACS), Siófok, Hungary*. 2012.
- RITCHIE, G. The jape riddle generator: technical specification. EDI-INF-RR 0158, School of Informatics, University of Edinburgh, 2003.
- SHANNON, C. A mathematical theory of communication. *The Bell System Technical Journal*, vol. 27, páginas 379–423, 1948.
- SHANNON, C. Prediction and entropy of printed english. *The Bell System Technical Journal*, vol. 30, páginas 50–64, 1951.
- SHUTOVA, E., VAN DE CRUYS, T. y KORHONEN, A. Unsupervised metaphor paraphrasing using a vector space model. En *Proceedings of the 24th International Conference on Computational Linguistics*, páginas 1121–1130. 2012.
- SHUTOVA, E., SUN, L. y KORHONEN, A. Metaphor identification using verb and noun clustering. En *Proceedings of the 23rd International Conference on Computational Linguistics*, páginas 1002–1010. Association for Computational Linguistics, 2010.
- STOCK, O. y STRAPPARAVA, C. Hahacronym: Humorous agents for humorous acronyms. *Humor: International Journal of Humor Research*, vol. 16(3), páginas 297–314, 2003.

- TINHOLT, H. y NIJHOLT, A. Computational humour: Utilizing cross-reference ambiguity for conversational jokes. En *Proceedings of International Workshop on Fuzzy Logic and Applications*. 2007.
- TURNER, S. Minstrel: a computer model of creativity and storytelling. *Association for Computational Linguistics*, 1993.
- VEALE, T. Creative language retrieval: A robust hybrid of information retrieval and linguistic creativity. En *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, páginas 278–287. 2011.
- VEALE, T. A service-oriented architecture for metaphor processing. En *Proceedings of the Second Workshop on Metaphor in NLP*, páginas 52–60. Association for Computational Linguistics, 2014.
- VEALE, T. y HAO, Y. Comprehending and generating apt metaphors: a web-driven, case-based approach to figurative language. En *Proceedings of the 22nd AAAI Conference*. 2007.
- VEALE, T. y HAO, Y. Exploiting readymades in linguistic creativity: A system demonstration of the jigsaw bard. En *ACL (System Demonstrations)*, páginas 14–19. 2011.
- VEALE, T. y LI, G. Specifying viewpoint and information need with affective metaphors: A system demonstration of the metaphor magnet web app/service. En *Proceedings of the ACL 2012 System Demonstrations*, páginas 7–12. 2012.
- VEALE, T. y LI, G. Creating similarity: Lateral thinking for vertical similarity judgments. En *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, páginas 660–670. 2013.
- WALLER, A., BLACK, R., O'MARA, D., PAIN, H., RITCHIE, G. y MANURUNG, R. Evaluating the standup pun generating software with children with cerebral palsy. *ACM Transactions on Accessible Computing*, vol. 1(3), páginas 16:1–16:27, 2009.
- WALLINGTON, A., BARNDEN, J., BUCHLOVSKY, P., FELLOWS, L. y GLASBEY, S. Metaphor annotation: A systematic study. CSRP 03-04, School of Computer Science, The University of Birmingham, 2003.
- WEINBERG, G., RAMAN, A. y MALLIKARJUNA, T. Interactive jamming with shimon: a social robotic musician. En *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, páginas 233–234. Association for Computational Linguistics, 2009.

WIKIPEDIA. Java api for restful web services. 2015. [Online; accessed 6-September-2017].

WIKIPEDIA. Concept map. 2017a. [Online; accessed 6-September-2017].

WIKIPEDIA. Ontology (information science). 2017b. [Online; accessed 6-September-2017].

WIKIPEDIA. Trope (literature). 2017c. [Online; accessed 6-September-2017].

WILKS, Y., GALESCU, L., ALLEN, J. y DALTON, A. Automatic metaphor detection using large-scale lexical resources and conventional metaphor extraction. En *Proceedings of the First Workshop on Metaphor in NLP*, páginas 36–44. Association for Computational Linguistics, 2013.

Apéndice A

Evaluación de acertijos

En las tablas A.1 y A.2 se presenta la lista completa de acertijos utilizados en la primera evaluación del sistema. La primera columna, *Acertijo*, muestra la posición del acertijo en la evaluación, y se refiere a los acertijos de las tablas 6.3 y 6.4. La segunda columna, *Palabra*, muestra el concepto de entrada del sistema, que debe ser acertado por los usuarios. En la tercera columna de la Tabla A.1, *Acertijo aleatorio*, se pueden ver las secuencias de pistas, en forma de acertijos, que son dados a los usuarios para adivinar el concepto buscado en la versión aleatoria de la evaluación. Por otro lado, en la tercera columna de la Tabla A.2, *Acertijo revisado*, se muestra las secuencias de pistas que son proporcionadas a los usuarios en la versión revisada de la evaluación.

| <i>Acertijo</i> | <i>Palabra</i> | <i>Acertijo aleatorio</i> |
|-----------------|------------------|--|
| R1 | <i>Hormiga</i> | ... diminuto como un isópodo? ... terrestre como un pájaro? ... pequeño como un conejo? ... social como una boda? |
| R2 | <i>Diablo</i> | ... inútil como la arena? ... común como el cromó? ... malvado como la envidia? ... invisible como la alegría? |
| R3 | <i>Sol</i> | ... estelar como un núcleo galáctico? ... caliente como una sopa? ... natural como la madera? ... gravitacional como un planeta? |
| R4 | <i>Coche</i> | ... grande como un caballo? ... físico como una dureza? ... privado como un hotel? ... técnico como un medicamento? |
| R5 | <i>Ballena</i> | ... marino como un percebe? ... grande como un mueble? ... migratorio como un ganso? ... acuático como un pez? |
| R6 | <i>Diamante</i> | ... transparente como un pelo? ... puro como el oro? ... costoso como un coche? ... simple como un destornillador? |
| R7 | <i>Leche</i> | ... líquido como el metanol? ... crudo como el algodón? ... natural como la madera? ... cotidiano como la ropa? |
| R8 | <i>Tiburón</i> | ... grande como un caballo? ... peligroso como unas tijeras? ... marino como un percebe? ... acuático como un pez? |
| R9 | <i>Coca cola</i> | ... comercial como un supermercado? ... duro como el hormigón? ... carbonatado como un refrigerador? ... fresco como el suelo húmedo? |
| R10 | <i>Avión</i> | ... físico como la natación? ... móvil como un camión? ... moderno como un restaurante? ... fijo como un árbol? |

Tabla A.1: Acertijos aleatorios utilizados en la evaluación

| <i>Acertijo</i> | <i>Palabra</i> | <i>Acertijo revisado</i> |
|-----------------|------------------|--|
| R1 | <i>Hormiga</i> | ... <i>invertebrado como un cangrejo?</i> ... <i>social como una boda?</i> ... <i>pequeño como un isópodo?</i> ... <i>molesto como una babosa?</i> |
| R2 | <i>Diablo</i> | ... <i>malvado como la envidia?</i> ... <i>sobrenatural como una deidad?</i> ... <i>poderoso como la ira extrema?</i> ... <i>malo como la guerra?</i> |
| R3 | <i>Sol</i> | ... <i>caliente como una sopa?</i> ... <i>estelar como un núcleo galáctico?</i> ... <i>amarillo como un mango?</i> ... <i>central como una sala de estar?</i> |
| R4 | <i>Coche</i> | ... <i>mecánico como un engranaje?</i> ... <i>cotidiano como la ropa?</i> ... <i>pesado como el plomo?</i> ... <i>grande como un caballo?</i> |
| R5 | <i>Ballena</i> | ... <i>grande como un mueble?</i> ... <i>migratorio como un ganso?</i> ... <i>marino como un percebe?</i> ... <i>acuático como un pez?</i> |
| R6 | <i>Diamante</i> | ... <i>duro como el cemento?</i> ... <i>transparente como un pelo?</i> ... <i>precioso como la plata?</i> ... <i>geométrico como un círculo?</i> |
| R7 | <i>Leche</i> | ... <i>blanco como un abadejo?</i> ... <i>líquido como el metanol?</i> ... <i>natural como la madera?</i> ... <i>crudo como el algodón?</i> |
| R8 | <i>Tiburón</i> | ... <i>peligroso como unas tijeras?</i> ... <i>marino como un percebe?</i> ... <i>depredador como un guepardo?</i> ... <i>grande como un tigre?</i> |
| R9 | <i>Coca cola</i> | ... <i>carbonatado como un refrigerador?</i> ... <i>comercial como un supermercado?</i> ... <i>oscuro como un higo?</i> ... <i>fresco como el suelo húmedo?</i> |
| R10 | <i>Avión</i> | ... <i>mecánico como una bomba?</i> ... <i>rápido como una hamburguesa?</i> ... <i>móvil como un camión?</i> ... <i>complicado como un barco?</i> |

Tabla A.2: Acertijos revisados utilizados en la evaluación

Apéndice B

Evaluación de figuras retóricas

La forma en que se crearon las analogías, símiles y metáforas de la evaluación de las figuras retóricas con las comparaciones desarrolladas por nuestro sistema fue la siguiente:

- Figuras retóricas comúnmente aceptadas: 6 sustantivos (3 abstractos y 3 concretos) se utilizaron como conceptos destino para obtener metáforas, símiles y analogías comúnmente aceptados:
 - TIME: Time is money / Time is like money / Time is as valuable as money
 - KNOWLEDGE: Knowledge is light / Knowledge is like light / Knowledge is as attractive as light
 - ARGUMENT: An argument is a war / An argument is like a war / An argument is as violent as a war
 - BALLERINA: A ballerina is a swan / A ballerina is like a swan / A ballerina is as graceful as a swan
 - STAR: A star is a diamond / A star is like a diamond / A star is as bright as a diamond
 - THUNDER: A thunder is a lion / A thunder is like a lion / A thunder is as mighty as a lion
- Figuras retóricas generadas aleatoriamente: 6 sustantivos (3 abstractos y 3 concretos) se utilizaron como conceptos destino para obtener metáforas, símiles y analogías generadas aleatoriamente:
 - HUNGER: Hunger is knowledge / Hunger is like knowledge / Hunger is as mechanical as knowledge

- SAILING: Sailing is boyhood / Sailing is like boyhood / Sailing is as allergenic as boyhood
 - SYLLOGISM: A syllogism is a nation / A syllogism is like a nation / A syllogism is as ungulate as a nation
 - ELEPHANT: An elephant is a napkin / An elephant is like a napkin / An elephant is as holy as a napkin
 - CORKSCREW: A corkscrew is a stamp / A corkscrew is like a stamp / A corkscrew is as furry as a stamp
 - TRAIN: A train is a violin / A train is like a violin / A train is as observational as a violin
- Figuras retóricas generadas automáticamente: 24 sustantivos (12 abstractos y 12 concretos) se utilizaron como conceptos destino por nuestro sistema para obtener metáforas, símiles y analogías. La mitad de ellos se generaron con el sistema configurado de modo que se obtuviese el concepto fuente de la misma categoría que el concepto destino, y la otra mitad de forma que se tomase el concepto fuente de una categoría diferente.
- Conceptos fuente y destino de la misma categoría:
 - WEDDING: A wedding is a party / A wedding is like a party / A wedding is as private as a party
 - WISH: A wish is a desire / A wish is like a desire / A wish is as mental as a desire
 - LIFE: Life is politics / Life is like politics / Life is as complex as politics
 - ANGEL: An angel is a fairy / An angel is like a fairy / An angel is as invisible as a fairy
 - DEVIL: Devil is love / Devil is like love / Devil is as spiritual as love
 - GOVERNMENT: Government is family / Government is like family / Government is as social as family
 - SNOW: Snow is a carpet / Snow is like a carpet / Snow is as soft as a carpet
 - NEEDLE: A needle is a knife / A needle is like a knife / A needle is as sharp as a knife
 - COTTON: Cotton is cashmere / Cotton is like cashmere / Cotton is as natural as cashmere

- HONEY: Honey is sugar / Honey is like sugar / Honey is as sticky as sugar
- BATTLE: A battle is a war / A battle is like a war / A battle is as historical as a war
- WRITER: A writer is a designer / A writer is like a designer / A writer is as creative as a designer
- Conceptos fuente y destino de diferentes categorías:
 - SAVING: Saving is farming / Saving is like farming / Saving is as productive as farming
 - ACCIDENT: An accident is an electric shock / An accident is like an electric shock / An accident is as unexpected as an electric shock
 - NETWORK: Network is family / Network is like family / Network is as social as family
 - IDEA: Idea is colors / Idea is like colors / Idea is as abstract as colors
 - ASSEMBLY: An assembly is an aircraft / An assembly is like an aircraft / An assembly is as complex as an aircraft
 - WINTER: Winter is salad / Winter is like salad / Winter is as cold as salad
 - MOON: The moon is an halogen lamp / The moon is like an halogen lamp / The moon is as bright as an halogen lamp
 - REFUGEE: A refugee is an elderly / A refugee is like an elderly / A refugee is as vulnerable as an elderly
 - TEMPLE: A temple is a school / A temple is like a school / A temple is as public as a school
 - ACID: Acid is a tiger / Acid is like a tiger / Acid is as dangerous as a tiger
 - BULLET: A bullet is a bolt / A bullet is like a bolt / A bullet is as metal as a bolt
 - DRAWER: A drawer is a chesnut / A drawer is like a chesnut / A drawer is as dark as a chesnut

Apéndice C

Artículo publicado para Language Resources and Evaluation Conference 2016

Riddle Generation using Word Associations

El siguiente artículo, acerca de la generación de adivinanzas, cuya publicación fue aceptada bajo el nombre *Riddle Generation using Word Associations*, se presentó como póster en el congreso *Language Resources and Evaluation Conference* en el año 2016. En el artículo se habla de la información y el conocimiento disponible útil para el desarrollo de aplicaciones informáticas y cómo ha sido utilizada en trabajos previos relacionados con la generación de adivinanzas y de herramientas de asociación de conceptos. Posteriormente, se explica cómo se utiliza la información obtenida de la herramienta Thesaurus Rex para implementar nuestro propio servicio generador de acertijos. Se describe la evaluación realizada con los acertijos de nuestro sistema y se discuten los resultados obtenidos para, por último, obtener conclusiones.

Riddle Generation using Word Associations

Paloma Galván¹, Virginia Francisco¹, Raquel Hervás¹, Gonzalo Méndez²

¹Departamento de Ingeniería del Software e Inteligencia Artificial

²Instituto de Tecnología del Conocimiento

Universidad Complutense de Madrid, Spain

palomagalvan@ucm.es, {virginia,raquelhb,gmendez}@fdi.ucm.es

Abstract

In knowledge bases where concepts have associated properties, there is a large amount of comparative information that is implicitly encoded in the values of the properties these concepts share. Although there have been previous approaches to generating riddles, none of them seem to take advantage of structured information stored in knowledge bases such as Thesaurus Rex, which organizes concepts according to the fine grained ad-hoc categories they are placed into by speakers in everyday language, along with associated properties or modifiers. Taking advantage of these shared properties, we have developed a riddle generator that creates riddles about concepts represented as common nouns. The base of these riddles are comparisons between the target concept and other entities that share some of its properties. In this paper, we describe the process we have followed to generate the riddles starting from the target concept and we show the results of the first evaluation we have carried out to test the quality of the resulting riddles.

Keywords: Computational Creativity, Riddle Generation, Word Associations

1. Introduction

In knowledge bases where concepts have associated properties, there is a large amount of comparative information that is implicitly encoded in the values of the properties that these concepts share. This kind of information can be useful in tasks where it is required to automatically establish relations between concepts, such as the generation of comparisons between entities based on shared properties (e.g. *this shirt is as white as snow*). These tasks are included in computational creativity, trying to simulate the natural behaviour of human beings to be creative using computer programs.

In this paper we present a riddle generator that creates riddles about concepts represented as common nouns. The base of these riddles are comparisons between the target concept (i.e. *a shirt*) and other entities that share some of its properties (i.e. *snow*). The resulting riddles are composed as a sequence of comparisons following this template: “What is... as *attribute* as *concept*?”, where *attribute* is a property of the target concept which is the answer to the riddle, and *concept* is a different entity that shares the value of the *attribute* with the target concept. For example, “What is... as hard as concrete and as transparent as hair?” is a riddle generated for the concept *diamond* by the riddle generator.

In order to gather information about the features that characterize the target concept of the riddle, and to obtain similar concepts according to those features, a word association resource called Thesaurus Rex (Veale and Li, 2013) has been used.

This paper is organized as follows. Section 2 provides an overview of the state of the art of riddle and puzzle generation. Section 3 explains our approach to generate riddles using word associations. In Section 4 we present the evaluation and results obtained from a series of questionnaires where people tried to solve two sets of riddles generated by our system. Finally, Section 5 contains some conclusions and future work.

2. Related Work

Although the generation of riddles may seem a difficult task from a computational point of view, there have been previous attempts to the automatic generation of riddles.

De Palma and Weiner (1992) propose a model of a knowledge representation that contains the data to generate or solve riddles. They develop an algorithm that generates a guess based on homophonous concepts.

JAPE (Binsted and Ritchie, 1997; Ritchie, 2003) is a computer program which generates simple punning riddles using templates with slots where words or phrases are inserted. To determine which words must to be incorporated to the final riddle, the system makes use of predefined schemas (manually built from previously known jokes), which establish relationships between words which must hold to build a joke. The program was tested by 120 children that rated generated riddles, human-generated texts, and non-joke texts for ‘jokiness’ and ‘funniness’. The evaluation confirmed that riddles generated were jokes, and that there is no significant difference in ‘funniness’ or ‘jokiness’ between punning riddles generated by their system and published human-generated jokes.

Some of the authors of JAPE (Cunningham et al., 2000) have furtherly developed STANDUP (Waller et al., 2009), a large-scale pun generator to allow children with communication disabilities to improve their linguistic skills. The pun generation followed the same steps used in JAPE, but several improvements had to be introduced in order to adapt the generated puns to the target audience, i.e. children with communication disabilities: speech output, picture support, restricted topics or use of familiar words. The system was evaluated with real users over a short period, and although no positive effects could be observed on the long term, the authors report a change in the attitude of the children towards communication.

Colton (2002a) extended the HR automated theory formation system (Colton, 2002b) to enable it to automatically generate puzzles given background information about a set

of objects of interest. They generate three types of puzzles: odd one out, next in sequence and analogy puzzles. They found that the main problem with puzzle generation was ensuring the uniqueness of the concept supposed to explain the puzzle solution.

Pintér et al. (2012) propose a knowledge-lean method to generate three types of word puzzles (odd one out, choose the related word, and separate the topics) from unstructured and unannotated document collections. The difficulty of the puzzles can be adjusted. The algorithm is based on topic models, semantic similarity, and network capacity.

Guerrero et al. (2015) present a Twitter bot that generates riddles about celebrities. The model selects a celebrity, retrieves relevant traits to describe him, generates analogies between his attributes and converts such descriptions into utterances, and, finally, tweets the generated riddle and interact with users by evaluating their answers. To evaluate the riddle generation they asked 86 people to evaluate five riddles. They first asked the participants to guess the answer to the riddle. Then, they presented the correct answer and asked if they knew the person in question. The participants indicated whether they considered the quality of the riddle satisfactory and, if negative, gave the reason why it was not good. The percentage of known celebrities once the answer was presented (54.19%) indicates that the process for the selection of celebrities should be improved. The low number of correct answers (15.58%) suggests that the complexity of the generated riddles was high.

3. Riddle Generation Using Word Associations

The proposed riddle generator receives a common noun as an input, which is the target concept for the riddle. Using Thesaurus Rex, a database of word associations extracted from the web, the system unfolds a series of comparisons between the target concept and other concepts with similar properties in order to create the final riddle.

3.1. Thesaurus Rex

Thesaurus Rex (Veale and Li, 2013) organizes concepts according to the fine grained ad-hoc categories they are placed into by speakers in everyday language (*food, drink, beverage...*). These categories have an associated weight that represents their relative importance for the given concept. Thesaurus Rex can show different categories for each concept and allows in turn to consult the concepts in each category. If we take as an example the concept *coffee*, some of its categories with more weight are *beverage* or *drink* and some with less weight are *seed* or *poison*. Table 1 shows some categories for coffee and their corresponding weights. Concepts in Thesaurus Rex have also associated properties or modifiers which are also accompanied by a weight indicating how strong its relation to the concept is. For example, for *coffee* some of the modifiers with more weight are *hot, acidic* or *stimulating*, and modifiers with less weight are *granulated* or *digestive*.

3.2. Riddle Generation

Table 2 shows a few examples of target concepts and how Thesaurus Rex is used to generate riddles. Taking the first

| Category | Weight | Attribute | Weight |
|------------|--------|---------------|--------|
| drink | 4983 | hot | 3900 |
| smell | 185 | granulated | 10 |
| beverage | 7056 | acidic | 2909 |
| seed | 3 | dark | 1144 |
| intoxicant | 14 | stimulating | 1267 |
| liquid | 2541 | noncarbonated | 24 |
| food | 3322 | colored | 696 |
| poison | 5 | digestive | 7 |
| ... | ... | ... | ... |

Table 1: Examples of categories and attributes (including weights) for the concept *coffee* on Thesaurus Rex

concept, *sun*, as an example, the detailed process to generate a riddle is the following:

1. **Target concept categories.** To obtain the filtered categories to which the target concept belongs, we first extract a list of all the general categories of the concept using a Thesaurus Rex query. From this list, only the $N\%$ of categories with the highest weights are considered as candidates. The value of N is configurable. If a high N value is set, we will have in the list categories with lower weights, which are less relevant to the target concept. In the same way, we can set N to a low value, facing the risk of shortening the list to a single element. In the *sun* example, the categories with higher weights in Thesaurus Rex are *body* and *object*.
2. **Modifier extraction.** In addition to the categories, we also need a list of modifiers associated to the target concept, which is returned by a new query to Thesaurus Rex. From this list, the $N\%$ of attributes with the highest weights are considered as candidates. For example, if our target concept is the noun *sun*, some of the most important properties extracted are: *stellar, hot, natural* and *yellow*.
3. **Modifier selection.** One of the modifiers previously obtained is randomly selected. This random selection makes the system less repetitive, as the riddles obtained for the same target concept are not always the same as if only the modifier with the highest weight were selected. For the current example, we suppose that the system has chosen the modifier *hot*.
4. **New categories selection.** Using the modifier chosen in the previous step, a new query to Thesaurus Rex is performed in order to obtain new categories that also present this modifier as a highlighted property. In order to obtain comparisons between different kinds of concepts, the new categories that match the categories obtained in step 1 are discarded. In this way, we are avoiding the comparison of the target concept with other concepts in the same category in order to obtain more creative results. In the *sun* example, the new categories selected could be *food* and *beverage*, which are categories that present the *hot* property in Thesaurus Rex.

| <i>Target concept</i> | <i>Categories</i> | <i>Modifiers</i> | <i>New categories for the selected modifier</i> | <i>New query</i> | <i>Obtained concepts new query</i> | <i>Comparison</i> |
|-----------------------|-----------------------------|--|---|------------------|------------------------------------|------------------------|
| sun | body, object, star... | stellar, hot , natural, yellow... | food , beverage ... | hot food | chili, soup , garlic... | as hot as soup? |
| whale | animal, mammal, predator... | large, migratory , marine, aquatic... | bird , fish... | migratory bird | goose , duck, heron... | as migratory as goose? |
| diamond | stone, material, gem... | precious, valuable, hard , crystalline... | material, surface ... | hard surface | wood, wall, concrete ... | as hard as concrete? |

Table 2: Examples of comparisons obtained to be part of the riddles. Words in bold represent the choices made for each example.

5. **New category selection.** One of the new categories obtained in the previous step is randomly selected. For the current example, *food* is supposed to be the category selected.
6. **New query composition.** A new query for Thesaurus Rex is then composed by using the new category obtained in the previous step and the modifier selected in step 3. In the current example, we will assume this new query is *hot food*.
7. **Final concept selection.** With the query composed in the previous step, we obtain a list of concepts that belong to the category selected in step 5 (*food*) and at the same time present the property selected in step 3 (*hot*). This list is usually quite extensive, so the system randomly chooses among the results that have an associated weight among the $N\%$ of concepts with the highest weights. In our example, a possible final concept for *hot food* is *soup*.
8. **Comparison template.** With the final data obtained during this process, attribute and new concept, the template “as *attribute* as *concept*?” is filled. The result of this round is “as *hot* as *soup*?”.
9. **Riddle composition.** Steps 3-8 are repeated as many times as desired, determined in the configuration of the system. In each round, a new comparison is generated and added to the final riddle. In our example, a possible riddle with three comparisons is the following:

*What is ...
... as hot as soup?
... as stellar as a galactic nucleus?
... as yellow as a mango?*

4. Evaluation

We have carried out an evaluation to test whether word associations obtained by our system provided useful information for riddle generation, and to assess the quality of the resulting riddles. In order to do that, human evaluators were asked to guess the initial concepts which were used to create the riddles. Then, we studied the rate of success obtained by the evaluators, while at the same time analyzing how many comparisons were required to obtain the correct answers in different riddles. Some issues related to ambiguity and contradiction appeared when creating the riddles, so

we decided to create two different sets of riddles to perform the evaluation.

4.1. Design

Ten riddles were presented to human evaluators to see if they were able to find the initial target concepts. Riddles were presented in four phases, in order to know how many comparisons were needed to solve the riddle. In the first phase a single comparison was presented, in the second phase two comparisons were presented, three comparisons in the third phase and, finally, four comparisons in the fourth phase. The evaluation was carried out using Google Forms and some personal information was collected for statistical purposes (age, gender and riddle ability).

As explained in the previous section, the comparisons used in riddles are randomly chosen. When generating riddles for the evaluation, we realized that some of the comparisons do not add new information to previous ones, or the information added was contradictory or not valid due to the polysemy of some concepts. This is the case of *coke*, which is a flavoured carbonated drink and the street name for cocaine. Our system had generated “*is as carbonated as ...*” and “*is as hard as ...*”. The second comparison is obtained due to the word association “*coke - hard drug*”. Examples of contradictory comparisons are mostly related to attributes with imprecise values, like size or age. For example, Thesaurus Rex categorizes the concept *dog* as both small and large depending on the context. However, if our system chooses both attributes, we would have contradictory comparisons in the riddle.

In order to carry out a more detailed evaluation, we decided to create two different riddle sets. Using the same ten concepts, but with some differences in the provided comparisons, we created an original and a curated version of the riddles. For the first set, the resulting comparisons were randomly selected. For the other set, the four most significant comparisons were manually selected among seven generated using the described process in order to avoid not valid comparisons due to polysemy or semantic contradictions. The riddles used in the evaluation can be seen in an Appendix at the end of the paper.

4.2. Results

Both the evaluation with the random riddles and the one with the curated versions were performed in parallel by 12 different evaluators each, making a total of 24 participants in the experiment. The order of appearance for each riddle

| Phases | R1 | R2 | R3 | R4 | R5 | R6 | R7 | R8 | R9 | R10 |
|---------|-----|----|-----|-----|-----|-----|----|-----|-----|-----|
| Phase 1 | 17% | 0% | 17% | 8% | 0% | 0% | 0% | 8% | 0% | 0% |
| Phase 2 | 8% | 0% | 67% | 0% | 33% | 25% | 8% | 17% | 0% | 0% |
| Phase 3 | 8% | 0% | 58% | 33% | 50% | 92% | 8% | 67% | 0% | 0% |
| Phase 4 | 8% | 0% | 42% | 17% | 50% | 67% | 0% | 75% | 17% | 0% |

Table 3: Percentage of success for each riddle in random set

| Phases | R1 | R2 | R3 | R4 | R5 | R6 | R7 | R8 | R9 | R10 |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Phase 1 | 0% | 25% | 0% | 17% | 0% | 8% | 17% | 0% | 33% | 0% |
| Phase 2 | 33% | 25% | 75% | 33% | 25% | 58% | 67% | 75% | 33% | 0% |
| Phase 3 | 42% | 25% | 75% | 42% | 58% | 83% | 75% | 75% | 67% | 8% |
| Phase 4 | 42% | 33% | 92% | 50% | 58% | 75% | 75% | 75% | 67% | 8% |

Table 4: Percentage of success for each riddle in curated set

| Phases | Random set | Curated set |
|---------|------------|-------------|
| Phase 1 | 5% | 10% |
| Phase 2 | 16% | 43% |
| Phase 3 | 32% | 55% |
| Phase 4 | 28% | 58% |

Table 5: Percentage of guessed riddles

was fixed, so all the evaluators participating in each part of the experiment were presented exactly the same riddles. To view the results for each riddle, Tables 3 and 4 show disaggregated percentages of success for each riddle in each phase. The percentages of correctly guessed riddles in the evaluation for each phase are presented in Table 5.

4.3. Discussion

The approach we have used to generate the riddles assures that all of them have a solution. In the random generation mode, all the participants of the evaluation guessed at least two riddles (5%) and thirteen at most (32.5%), with an average of eight riddles guessed (20%) per person. In the curated generation mode, the minimum amount of guessed riddles was 6 (15%) and the maximum was 24 (60%), with an average of 16.5 riddles guessed (41.3%) per person. The aim is not to get all the riddles solved, which would indicate that they are too easy.

As shown in Table 5 the results of the curated version of the riddles are significantly better than the ones of the random version. So, it is evident that a special selection of comparisons is needed in some cases.

Regarding the number of comparisons needed to guess the correct answer, a curious fact can be seen in Table 5. With just a single comparison, there is almost no chance of guessing the target concept. In most cases, people answer at random because there are lots of concepts that share the presented attribute. When providing two comparisons, users are able to multiply by four the number of correct guesses. When they are provided three comparisons, in the case of randomly chosen comparisons, they reach their maximum rate of success. In the case of manually selected compar-

isons, they guess 55% of riddles, which is almost the maximum success, because the difference with the last phase, where four comparisons are provided, is almost negligible. At a more detailed level, Tables 3 and 4 shows that R10 has, in the best case scenario, a success rate of 8%. This is due to the fact that the attributes selected are not specific enough and there is a large amount of common properties with other concepts. In this example, the concept was *aircraft* and the attributes selected were: *mechanical*, *fast*, *mobile* and *complicated*. However, the third concept, *sun*, in the fourth phase of the curated set, had a success rate of 92%, as shown in Table 4. The reason for this is that the attributes selected for this concept were much more specific. For instance, an attribute that not many concepts share is *stellar*, which combined with *yellow*, *hot* and *central* limits the possible answers for this riddle.

From the point of view of the success rate in each set, in the last phase of the random set (Table 3), sometimes the percentage of correct answers decreases slightly. The reason for this, as explained by the participants in the evaluation, is that sometimes the last hints were contradictory, and users were confused and ended up changing their answer in the last attempt. However, in the curated set (Table 4), only R6 presents a decrease of the success rate in the last phase. This means that the last comparison in this case was confusing for some of the evaluators.

5. Conclusions and Future Work

An automated mechanism for riddle generation using Thesaurus Rex, a resource based on word associations, has been presented. Following the described process to generate the riddles, the subsequent evaluation points out that the word associations obtained by our system are useful for generating these riddles. However, the evaluation also shows that a manual selection of comparisons is useful because confusing comparisons may be generated when the target of the riddles is a polysemic concept or presents some contradictory attributes. Therefore, it is necessary to develop some mechanisms to select only the modifiers related to the sought meaning of the target concept, and consider in a special way attributes with imprecise values.

The results of the evaluation also suggest that the order in which the comparisons are provided is relevant in order to solve the riddle using less comparisons, so it may be useful to analyze the discriminating power of each attribute, so that the complexity of the riddles can be controlled. If this information is available, the system could select first (or last) the most discriminating attributes of the concept automatically. The underlying idea is that the higher the discriminating power of the attribute, the easier the riddle, as more concepts are excluded from the possible answers. Depending on the desired difficulty of the riddle, we can play with the order of the attributes according to their discriminating power.

As seen during evaluation, to make better guesses three or more comparisons are generally needed. In the future, we will evaluate with five or more comparisons to determine an optimal number in order to have riddles that are not impossible to guess, while at the same time are not too obvious. In addition, we would like to include riddles created by humans in future evaluations, so we can assess whether our riddles are easier or more difficult in comparison, and if they are considered natural in comparison with human-made ones.

In the future, we will explore the possibility of developing more creative riddles, for example with rhymes and a more elaborated selection of attributes and concepts.

One of our main concerns when developing the described riddle generator is the practical application for the resulting system. Currently, we are using a similar approach to generate rhetorical figures, such as analogies, similes and metaphors. In this case, instead of hiding the target concept and making the user guess it, the similarities between the two concepts are explored in order to create tropes that are as evocative and meaningful as possible. Hence, one of our current goals with riddle generation is to study the relationship between concepts through shared properties to gain a deeper insight that helps us generate better linguistic resources.

Another research line where we are starting to explore the applicability of the described techniques is accessibility, and more specifically, text simplification and text generation for users with cognitive disabilities. Other authors have already reported on the use of riddles to allow children with communication difficulties to develop their linguistic skills (Manurung et al., 2008). Following this idea, we aim at exploring the way in which riddles can be incorporated in the life of people with communication disabilities, supported, in addition, with the use of pictographs.

6. Acknowledgements

This work is funded by ConCreTe. The project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 611733.

7. References

Binsted, K. and Ritchie, G. (1997). Computational rules for generating punning riddles. *Humor - International Journal of Humor Research*, 10(1):25–76.

- Colton, S. (2002a). Automated puzzle generation. In *Proceedings of the AISB'02 (Symposium on AI and Creativity in the Arts and Science)*.
- Colton, S. (2002b). *Automated Theory Formation in Pure Mathematics*. Springer series on distinguished dissertations.
- Cunningham, H., Maynard, D., and Tablan, V. (2000). JAPE: a Java Annotation Patterns Engine (Second Edition). Research Memorandum CS-00-10, Department of Computer Science, University of Sheffield, November.
- De Palma, P. and Weiner, E. (1992). Riddles: accessibility and knowledge representation. In *Proceedings of Coling'92*, pages 1121–1125.
- Guerrero, I., Verhoeven, B., Barbieri, F., Martins, P., and Pérez y Pérez, R. (2015). Theriddlerbot: A next step on the ladder towards creative twitter bots. In *Proceedings of the Sixth International Conference on Computational Creativity*, pages 315–322.
- Manurung, R., Ritchie, G., Pain, H., Waller, A., O'Mara, D., and Black, R. (2008). The construction of a pun generator for language skills development. *Applied Artificial Intelligence*, 22(9):841–869.
- Pintér, B., Voros, G., Szabo, Z., and Lorincz, A. (2012). Automated word puzzle generation using topic models and semantic relatedness measures. In *Joint Conference on Mathematics and Computer Science (MACS)*, Siófok, Hungary.
- Ritchie, G. (2003). The jape riddle generator: technical specification. EDI-INF-RR 0158, School of Informatics, University of Edinburgh.
- Veale, T. and Li, G. (2013). Creating similarity: Lateral thinking for vertical similarity judgments. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 660–670.
- Waller, A., Black, R., O'Mara, D. A., Pain, H., Ritchie, G., and Manurung, R. (2009). Evaluating the standup pun generating software with children with cerebral palsy. *ACM Transactions on Accessible Computing*, 1(3):16:1–16:27.

Appendix: Evaluation Riddles

The complete list of riddles used in the evaluation is presented in Table 6. The first column, *Riddle*, shows the position of the riddle in the evaluation, and refers to the riddles shown in Tables 3 and 4. The second column, *Word*, shows the target concept that had to be guessed by the users. In the third column, *Random riddle*, we show the sequence of clues, in the form of riddles, that were given to the users to guess the target word in the random version of the evaluation. Finally, in the last column, *Curated riddle*, we show the sequence of clues that were provided to the users in the curated version of the evaluation.

| <i>Riddle</i> | <i>Word</i> | <i>Random riddle</i> | <i>Curated riddle</i> |
|---------------|-----------------|--|---|
| R1 | <i>Ant</i> | ... as tiny as an isopod? ... as terrestrial as a bird? ... as small as a rabbit? ... as social as a wedding? | ... as invertebrate as a crab? ... as social as a wedding? ... as tiny as an isopod? ... as annoying as a slug? |
| R2 | <i>Devil</i> | ... as useless as sand? ... as common as chromium? ... as evil as envy? ... as invisible as joy? | ... as evil as envy? ... as supernatural as a deity? ... as powerful as extreme anger? ... as bad as a war? |
| R3 | <i>Sun</i> | ... as stellar as a galactic nucleus? ... as hot as a soup? ... as natural as wood? ... as gravitational as a planet? | ... as hot as a soup? ... as stellar as a galactic nucleus? ... as yellow as a mango? ... as central as a living-room? |
| R4 | <i>Car</i> | ... as large as a horse? ... as physical as a hardness? ... as private as a hotel? ... as technical as a medicine? | ... as mechanical as a gear? ... as everyday as clothing? ... as heavy as lead? ... as large as a horse? |
| R5 | <i>Whale</i> | ... as marine as a barnacle? ... as large as a furniture? ... as migratory as a goose? ... as aquatic as a fish? | ... as large as a furniture? ... as migratory as a goose? ... as marine as a barnacle? ... as aquatic as a fish? |
| R6 | <i>Diamond</i> | ... as transparent as a hair? ... as pure as gold? ... as costly as a car? ... as simple as a screwdriver? | ... as hard as concrete? ... as transparent as a hair? ... as precious as silver? ... as geometric as a circle? |
| R7 | <i>Milk</i> | ... as liquid as methanol? ... as raw as cotton? ... as natural as wood? ... as everyday as clothing? | ... as white as a pollock? ... as liquid as methanol? ... as natural as wood? ... as raw as cotton? |
| R8 | <i>Shark</i> | ... as large as a horse? ... as dangerous as scissors? ... as marine as a barnacle? ... as aquatic as a fish? | ... as dangerous as scissors? ... as marine as a barnacle? ... as predatory as a cheetah? ... as big as a tiger? |
| R9 | <i>Coke</i> | ... as commercial as a supermarket? ... as hard as concrete? ... as carbonated as a cooler? ... as cool as damp soil? | ... as carbonated as a cooler? ... as commercial as a supermarket? ... as dark as a fig? ... as cool as damp soil? |
| R10 | <i>Aircraft</i> | ... as physical as swimming? ... as mobile as a truck? ... as modern as a restaurant? ... as fixed as a tree? | ... as mechanical as a pump? ... as fast as a hamburger? ... as mobile as a truck? ... as complicated as a ship? |

Table 6: Riddles used in the evaluation

Apéndice D

Artículo publicado para International Conference on Computational Creativity 2016

Exploring the Role of Word Associations in the Construction of Rhetorical Figures

A continuación, se presenta el artículo redactado sobre la generación de figuras retóricas, cuya publicación fue aceptada bajo el nombre *Exploring the Role of Word Associations in the Construction of Rhetorical Figures* y fue presentado como póster en el congreso *International Conference on Computational Creativity* en el año 2016. En el artículo se introduce el concepto de lenguaje figurativo y las figuras retóricas metáfora, símil y analogía, que son generadas por nuestro servicio. Se hace un repaso de los sistemas que se aproximan a la generación de figuras retóricas y se presenta nuestro proceso generador de figuras retóricas que utiliza Thesaurus Rex para encontrar asociaciones de conceptos. Se realiza una descripción de la evaluación de las figuras generadas por nuestra herramienta junto con un par de conjuntos más para establecer líneas base. Finalmente, se obtienen conclusiones en base a la discusión de los resultados de la evaluación.

Exploring the Role of Word Associations in the Construction of Rhetorical Figures

Paloma Galván¹, Virginia Francisco¹, Raquel Hervás¹, Gonzalo Méndez², Pablo Gervás²

¹Departamento de Ingeniería del Software e Inteligencia Artificial

²Instituto de Tecnología del Conocimiento

Universidad Complutense de Madrid, Spain

palomagalvan@ucm.es, {virginia, raquelhb, gmendez}@fdi.ucm.es, pgervas@ucm.es

Abstract

Figurative language is a fundamental characteristic of elaborate forms of linguistic communication. We currently have very poor models of how figurative language may be constructed in computational terms. The overall aim of this paper is to identify possible regularities, intuitions or heuristics that may at a later stage be employed to drive a text generator that is capable of using this type of rhetorical figure.

Introduction

The use of figurative language is a fundamental tool in linguistic communication. One of the most easily identifiable characteristics of computer generated text is the tendency to stick to literal meanings. This is partly because literal meanings are unambiguous and have less risk of misinterpretation. But it is also in part due to the fact that we currently have very poor models of how figurative language may be constructed in computational terms. This paper explores the relationship between word associations as modelled in already available computational resources and the type of rhetorical figures that people employ regularly. The aim is to identify possible regularities, intuitions or heuristics that may at a later stage be employed to drive a text generator that is capable of using this type of rhetorical figure.

We consider three types of rhetorical figures or tropes. A *metaphor* is a widely-used literary mechanism which allows comparison between two disparate concepts. Metaphors transfer the qualities of one word to another, as in *Booger was a lion in the electoral arena*. Here, the qualities of *lion* (the source) are transferred to Booger (the target). A *simile* is a pointed, direct and explicit metaphor where two different things are compared to evolve a new meaning. A simile denotes the target to be like the source, and as such the target cannot totally be substituted by the source. A simile is a kind of metaphor where the comparison is made using the words “as” or “like”. For example, *Booger was like a lion*. An *analogy* links two disparate concepts by common properties, as in *Booger was as brave as a lion*. Here, the quality of being *brave* (the property) is used to link *lion* (the source) to *Booger* (the target).

Metaphors play an important role in communication, occurring as often as every third sentence (Shutova et al.

2012), so the generation of metaphors is essential for Natural Language Generation. The same occurs for analogies and similes.

Black (1955) made explicit that metaphors depend upon conceptual connections between networks of concepts. Inherent in this approach is the idea that metaphors are a matter of cross-domain mapping (Lakoff 1993). A metaphor is a cognitive process that builds or maps connections between networks of concepts as it occurs with similes and analogies. In consequence, to generate metaphors a conceptual structure is needed where every concept is placed not only taking into account its conventional usage but its diverse and unconventional usages (Veale 2014b). The best place to find this complex structure is the Web and that is where we are going to look for word associations in order to create our analogies, similes and metaphors.

This paper presents a new approach to finding word associations in the web using Thesaurus Rex (Veale and Li 2013). The potential of this system will be studied for the automatic generation of analogies, similes and metaphors.

This paper is organized as follows. The second section presents prior work on analogy, simile and metaphor generation. The third section explains our approach to finding word associations. In the fourth section the evaluation of our approach for rhetorical figures generation is presented. And finally, in the last section conclusions and future work are explained.

Related Work

Rhetorical figures have been the target of researchers in computational approaches to linguistics on and off for many years. However, only in recent years has the combination of available knowledge resources and accumulated insights allowed for the field to flourish. Metaphors have been widely studied in Natural Language Analysis but not so much in Natural Language Generation (NLG). There is a lot of work related to metaphor detection (Wilks et al. 2013), identification (Shutova, Sun, and Korhonen 2010), extraction and annotation (Wallington et al. 2003) but few related to metaphor generation. The reason can be that metaphor generation is as challenging as human creativity will allow. In this section the most important approaches for simile and metaphor generation are presented.

Approaches to Rhetorical Figures in NLG

In the field of natural language generation there have been a number of attempts to establish procedures for constructing rhetorical figures as important ingredients of generated spans of text. This has been attempted both in general terms (Hervás et al. 2006b) for different types of rhetorical figures, and for specific cases like analogies (Hervás et al. 2006a) or metaphors (Hervás et al. 2007). These attempts were all carried out before adequate sources of machine-readable knowledge were available and consequently suffered from a thirst of appropriate knowledge. The attempts considered the problem of rhetorical figure employment in text generation in general theoretical terms but lacked sufficient volume of explicit knowledge on the underlying semantics of words to be capable of practical generation.

Approaches to Conceptual Construction of Rhetorical Figures

The recent development of sources of knowledge that allow easy mining of large corpora of text for significant word associations has led to the emergence of a number of systems that rely on these for constructing rhetorical figures of different types.

Jigsaw Bard Jigsaw Bard (Veale and Hao 2011) is a web service that exploits linguistic readymades to generate similes on demand. Jigsaw Bard scans Google n-grams to index potential readymades which are then re-purposed as a simile. For example, given the adjectival property *quiet* Jigsaw Bard returns the simile “*The peaceful life of a monastery*”. The Jigsaw Bard is best understood as a creative thesaurus: for any given property (or blend of properties) selected by the user, the Bard presents a range of apt similes, and users must decide which similes are most suited to their descriptive purposes.

Thesaurus Rex Thesaurus Rex (Veale and Li 2013) is a web service that given two concepts (for example, War and Divorce) returns a phase cloud of the nuanced categories that are shared by both concepts (in the given example it returns a cloud that contains *traumatic-event*, *stressful-event*, *unexpected-event*...).

Thesaurus Rex organizes concepts according to categories they are placed into by speakers in everyday language (*food*, *drink*, *beverage*...). These categories have an associated weight that represents their relative importance for the given concept. Thesaurus Rex can show different categories for each concept and allows in turn to consult the concepts in each category. For example for the concept *coffee*, some of its categories with more weight are *beverage* or *drink* and some with less weight are *leaf* or *opposition*. Concepts in Thesaurus Rex have associated properties or modifiers which are accompanied by a non-standard weight indicating how strong its relation to the concept is. For example, for *coffee* some of the modifiers with more weight are *hot*, *acidic* or *stimulating*, and modifiers with less weight are *smaller* or *adult*.

Metaphor Magnet Metaphor Magnet (Veale and Li 2012) is a Web service that allows users to enter queries with sin-

gle terms (such as *leader*), compound terms with an affective spin (such as *good leader* or *+leader*), or copula statements (such as “*Steve Jobs is a +leader*”). For each input, the service marries its extensive knowledge of lexicalized stereotypes to the grand scale of the Google n-grams to generate the most appropriate affective elaborations and interpretations. In each case, Metaphor Magnet provides an explanation of its outputs. If *Steve Jobs* were to be viewed as a *master*, the properties *skilled*, *enlightened*, *free* and *demanding* are all highlighted as being most appropriate. Metaphor Magnet sees metaphor interpretation as a question of which properties are mapped from the source to the target.

Metaphor Magnet lacks a proposition level view of the world, in which stereotypes are linked to other stereotypes by arbitrary relations.

Metaphor Eyes Metaphor Eyes (Veale 2014a) employs a propositional model of the world that reasons with subject-relation-object triples rather than subject-attribute pairs (as Metaphor Magnet does). Metaphor Eyes acquires its world-model from a variety of sources and it views metaphor as a representational lever, allowing it to fill the holes in its weak understanding of one concept by importing relevant knowledge from a neighboring concept.

Metaphor Eyes metaphorize one concept (the source) as other concept (the target). Given *Scientist* and *Artist* it generates metaphors as “*Scientists develop ideas like artists*”.

Figure8 Figure8 (Harmon 2015) is a system that contains an underlying model for what defines creative and figurative comparisons, and evaluates its own output based on these rules. The system is provided with a model of the current world and an entity in the world to be described. A suitable vehicle is selected from the knowledge base, and the comparison between the two nouns is clarified by obtaining an understanding via corpora search of what these nouns can do and how they can be described. Sentence completion occurs by intelligent adaptation of a case library of valid grammar constructions. Finally, the comparison is ranked by the system based on semantic, prosodic, and knowledge-based qualities.

Word Association Generation

This section presents the proposed approach for the generation of word associations, which has been implemented as a web service. This service receives a common noun as an input, which is the target concept for the word association. Following the steps described in the Process section below, the system generates source concepts with similar properties to the target concept creating word associations.

Entry

The proposed approach receives a common noun as an input, which is the target concept for which the word association must be generated. Using Thesaurus Rex, the system unfolds a comparison between the target concept and another concept that acts as the source of the rhetorical figure, with similar properties to the target concept in order to create a word association.

Table 1: Examples of word associations obtained. Words in bold represent the choices made for each example.

| Step | Target | snow | thunder | network |
|------|---|---|---|---|
| 1 | Categories | surface, elements, weather... | noise, sound, event... | system, structure, entity... |
| 2 | Modifiers | natural, reflective, slippery , soft, white... | natural, loud , sudden, weather... | social , complex, adaptive, physical... |
| 3 | Categories for the selected modifier | surface , ground, stuff... | instrument , thing... | institution , event activity, science... |
| 4 | New query | slippery surface | loud instrument | social institution |
| 5 | Obtained concepts | satin, silk, nylon, polyester... | trumpet, drum, horn, saxophone... | family, government, religion... |

Process

Table 1 shows a few examples of target concepts and how Thesaurus Rex is used to obtain words associated to the target concepts. Taking the first concept, *snow*, as an example, the detailed process is the following:

1. **Target concept categories.** To obtain the filtered categories to which the target concept belongs, we first extract a list of all the general categories of the concept using a Thesaurus Rex query. From this list, only the $N\%$ of categories with the highest weights are considered as candidates. The value of N is configurable (in this example, $N = 0.4$). If a high N value is set, we will have in the list categories with lower weights, which are less relevant to the target concept. In the same way, we can set N to a low value, facing the risk of shortening the list to a single element. In the *snow* example, the categories with higher weights in Thesaurus Rex are *surface* and *weather*.
2. **Modifier extraction.** In addition to the categories, we also need a list of modifiers associated to the target concept, which is returned by a new query to Thesaurus Rex. From this list, the $N\%$ of attributes with the highest weights are considered as candidates (in this example, $N = 0.6$). For example, if our target concept is the noun *snow*, some of the most important properties extracted are: *natural*, *reflective*, *slippery*, *soft* and *white*.
Modifier selection. One of the modifiers previously obtained is randomly selected. This random selection makes the system less repetitive, as the words associated to the same target concept are not always the same as if only the modifier with the highest weight were selected. For the current example, the system has chosen the modifier *slippery*.
3. **Categories selection.** Using the modifier chosen in the previous step, a new query to Thesaurus Rex is performed in order to obtain categories that present this modifier as a highlighted property. In the *snow* example, the categories selected could be *surface*, *ground* and *stuff* which are categories that present the *slippery* property in Thesaurus Rex.
Category selection. One of the categories obtained in the previous step is selected. The system could be parametrized to select a category which contains the target concept (a category that matches one obtained in step 1). It could also be parametrized to choose a category in which

the target concept is not included (discarding categories that match those obtained in step 1). For the current example, *surface* is the selected category.

4. **New query composition.** A new query for Thesaurus Rex is then composed by using the category obtained in the previous step and the modifier selected in step 3. In the current example, this new query is *slippery surface*.
5. **Final concept selection.** With the query composed in the previous step, we obtain a list of concepts that belong to the category selected in step 5 (*surface*) and at the same time present the property selected in step 3 (*slippery*). This list is usually quite extensive, so the system randomly chooses among the results that have an associated weight among the $N\%$ of concepts with the highest weights (in this example, $N = 0.1$). In our example, the final concepts associated to the target concept are *satin*, *silk*, *nylon* or *polyester*.

Output

The system output is the source concept that gives rise to the rhetorical figure, related through a shared property with the original target concept provided by the user. The shared property is significant in both concepts, which means that the property has a high weight for both of them. The resulting source concept is randomly chosen from the list of generated concepts, and it is subsequently used to create a rhetorical figure.

Evaluation

The aim of this evaluation has been twofold. On the one hand, we intended to test the appropriateness of the analogies, similes and metaphors generated by our system, in order for us to be able to refine the process followed to generate them. On the other hand, we also expected to find out what kind of rhetorical figure is more enlightening for the evaluators and which one is closer to a rhetorical figure generated by humans.

Rhetorical Figures Generation using Word Associations

This approach uses the simplest and purest copula form for analogies, similes and metaphors:

- Analogy: *TARGET is as PROP as SOURCE*.
- Simile: *TARGET is like SOURCE*.

- Metaphor: *TARGET is SOURCE*.

Design of the Evaluation

The evaluation set was composed by 36 analogies, 36 similes and 36 metaphors. To create these elements, 36 different words were used as target concepts and one analogy, one simile and one metaphor were created for each of them. In order to avoid the possibility that one evaluator could evaluate several rhetorical figures related to the same target concept, the original data set was divided in three different subsets of 36 rhetorical figures. Each subset had 12 metaphors, 12 similes and 12 analogies, all of them created from a different target concept.

The evaluation was carried out as an online survey using Google Forms, where each evaluator received a link to one of the three surveys and was asked to score each of the figures using a Likert scale. Evaluators were asked to rate how appropriate or natural sounding each trope was, giving them a score from 1 to 7 (where 1 symbolizes a completely inappropriate trope and 7 represents a completely natural sounding trope). We chose to use the median and the mode because when working with the Likert scale, these are the most interesting metrics. Interpreting the average when managing categories such as "totally meaningful" or "totally meaningless", would not provide useful information. Adding the "totally meaningful" value (5) to two "meaningless" values (2) would result in an average of 4, but that is not a very rich interpretation. Traditional statisticians do not recommend using the average of the data when dealing with values in a Likert scale, which offers ordinal values.

In order to have two different baselines in our experiment to measure the quality of the figures generated by our system, we have used a set of commonly accepted rhetorical figures, together with a set of random manually generated ones, to compare them against the ones generated by our system.

The way in which the analogies, similes and metaphors were created was the following:

- Commonly accepted figures: 6 words (3 abstract and 3 concrete) were used as target concepts to obtain commonly accepted metaphors, similes and analogies:
 - TIME: Time is money / Time is like money / Time is as valuable as money
 - KNOWLEDGE: Knowledge is light / Knowledge is like light / Knowledge is as attractive as light
 - ARGUMENT: An argument is a war / An argument is like a war / An argument is as violent as a war
 - BALLERINA: A ballerina is a swan / A ballerina is like a swan / A ballerina is as graceful as a swan
 - STAR: A star is a diamond / A star is like a diamond / A star is as bright as a diamond
 - THUNDER: A thunder is a lion / A thunder is like a lion / A thunder is as mighty as a lion
- Randomly generated figures: 6 words (3 abstract and 3 concrete) were used as target concepts to obtain randomly generated metaphors, similes and analogies:

- HUNGER: Hunger is knowledge / Hunger is like knowledge / Hunger is as mechanical as knowledge
- SAILING: Sailing is boyhood / Sailing is like boyhood / Sailing is as allergenic as boyhood
- SYLLOGISM: A syllogism is a nation / A syllogism is like a nation / A syllogism is as ungulate as a nation
- ELEPHANT: An elephant is a napkin / An elephant is like a napkin / An elephant is as holy as a napkin
- CORKSCREW: A corkscrew is a stamp / A corkscrew is like a stamp / A corkscrew is as furry as a stamp
- TRAIN: A train is a violin / A train is like a violin / A train is as observational as a violin
- Automatically generated figures: 24 words (12 abstract and 12 concrete) were used as target concepts by our system to obtain metaphors, similes and analogies. Half of them were generated with the system configured to obtain the source concept from the same category as the target, and the other half to take the source concept from a different category.
 - Source and target from the same category:
 - * WEDDING: A wedding is a party / A wedding is like a party / A wedding is as private as a party
 - * WISH: A wish is a desire / A wish is like a desire / A wish is as mental as a desire
 - * LIFE: Life is politics / Life is like politics / Life is as complex as politics
 - * ANGEL: An angel is a fairy / An angel is like a fairy / An angel is as invisible as a fairy
 - * DEVIL: Devil is love / Devil is like love / Devil is as spiritual as love
 - * GOVERNMENT: Government is family / Government is like family / Government is as social as family
 - * SNOW: Snow is a carpet / Snow is like a carpet / Snow is as soft as a carpet
 - * NEEDLE: A needle is a knife / A needle is like a knife / A needle is as sharp as a knife
 - * COTTON: Cotton is cashmere / Cotton is like cashmere / Cotton is as natural as cashmere
 - * HONEY: Honey is sugar / Honey is like sugar / Honey is as sticky as sugar
 - * BATTLE: A battle is a war / A battle is like a war / A battle is as historical as a war
 - * WRITER: A writer is a designer / A writer is like a designer / A writer is as creative as a designer
 - Source and target from different categories:
 - * SAVING: Saving is farming / Saving is like farming / Saving is as productive as farming
 - * ACCIDENT: An accident is an electric shock / An accident is like an electric shock / An accident is as unexpected as an electric shock
 - * NETWORK: Network is family / Network is like family / Network is as social as family
 - * IDEA: Idea is colors / Idea is like colors / Idea is as abstract as colors

Table 2: Metaphor results.

| Source | Mode | | | Median | | |
|---------------------------------------|----------|----------|----------|----------|----------|----------|
| | Abstract | Concrete | Total | Abstract | Concrete | Total |
| Random | 1 | 1 | 1 | 1 | 1 | 1 |
| Commonly accepted | 7 | 7 | 7 | 6 | 5 | 5 |
| Generated (different category) | 1 | 1 | 1 | 2 | 2 | 2 |
| Generated (same category) | 7 | 1 | 7 | 5 | 4 | 5 |
| Generated | 1 | 1 | 1 | 3 | 3 | 3 |

Table 3: Simile results.

| Source | Mode | | | Median | | |
|---------------------------------------|----------|----------|----------|----------|----------|----------|
| | Abstract | Concrete | Total | Abstract | Concrete | Total |
| Random | 1 | 1 | 1 | 2 | 1 | 1 |
| Commonly accepted | 7 | 5 | 7 | 6 | 5 | 5 |
| Generated (different category) | 1 | 1 | 1 | 2 | 3 | 3 |
| Generated (same category) | 7 | 6 | 6 | 5 | 4 | 5 |
| Generated | 1 | 1 | 1 | 4 | 3 | 4 |

Table 4: Analogy results.

| Source | Mode | | | Median | | |
|---------------------------------------|----------|----------|----------|----------|----------|----------|
| | Abstract | Concrete | Total | Abstract | Concrete | Total |
| Random | 1 | 1 | 1 | 1 | 1 | 1 |
| Commonly accepted | 7 | 7 | 7 | 6 | 6 | 6 |
| Generated (different category) | 1 | 7 | 2 | 3 | 4 | 4 |
| Generated (same category) | 5 | 7 | 7 | 4 | 4 | 4 |
| Generated | 1 | 7 | 7 | 4 | 4 | 4 |

Table 5: General results of the evaluation.

| Source | Mode | | | Median | | |
|---------------------------------------|----------|----------|----------|----------|----------|----------|
| | Abstract | Concrete | Total | Abstract | Concrete | Total |
| Random | 1 | 1 | 1 | 1 | 1 | 1 |
| Commonly accepted | 7 | 7 | 7 | 6 | 5 | 6 |
| Generated (different category) | 1 | 1 | 1 | 2 | 3 | 3 |
| Generated (same category) | 7 | 7 | 7 | 5 | 4 | 5 |
| Generated | 1 | 1 | 1 | 4 | 4 | 4 |

- * ASSEMBLY: An assembly is an aircraft / An assembly is like an aircraft / An assembly is as complex as an aircraft
- * WINTER: Winter is salad / Winter is like salad / Winter is as cold as salad
- * MOON: The moon is an halogen lamp / The moon is like an halogen lamp / The moon is as bright as an halogen lamp
- * REFUGEE: A refugee is an elderly / A refugee is like an elderly / A refugee is as vulnerable as an elderly
- * TEMPLE: A temple is a school / A temple is like a school / A temple is as public as a school
- * ACID: Acid is a tiger / Acid is like a tiger / Acid is as dangerous as a tiger
- * BULLET: A bullet is a bolt / A bullet is like a bolt / A bullet is as metal as a bolt
- * DRAWER: A drawer is a chesnut / A drawer is like a

chesnut / A drawer is as dark as a chesnut

Results of the Evaluation

The evaluation was carried out by 72 evaluators, so that each of the 3 subsets of rhetorical figures was assessed by 24 different evaluators.

The evaluation results for the metaphors are shown in Table 2. Overall the results obtained from the evaluation were as expected, random tropes turned out to be the ones with lower ratings, with a median of 1, and tropes with higher ratings were commonly accepted ones, with a median of 6. Regardless of the type of rhetorical figure and whether they represent specific or abstract concepts, the median of these figures is 4 (3 for those of different categories and 5 for those belonging to the same category).

Interestingly, the modes are the same for abstract and concrete concepts tropes, regardless of how they were generated. The mode of both the random tropes and the tropes

generated by our system with different categories is 1, and the mode of the commonly accepted ones and the tropes generated in the same category also matches, with a value of 7.

If we take a closer look at the data subsets of the metaphors, similes and analogies, we can observe that the value of the medians of all figures generated randomly in the three data sets is 1. The results are more satisfactory for the commonly accepted figures, with median values between 5 and 6, proving that the evaluators did not take risks awarding the maximum score.

When we continue to analyze the subsets, we can see that the results obtained for the tropes belonging to different categories are less promising than those obtained for tropes with the same category, with variations between 2 and 4. In the case of the analogies, the median is the same for the ones generated in the same category or in different categories, with a value of 4. The difference between medians of random tropes and commonly accepted tropes fluctuates between 4 and 5.

The graphs show comparative results for the different ways of generating the rhetorical figures: using concrete and abstract concepts, as well as the combined results. The first graph (see Figure 1), corresponds to the word association using abstract concepts and we can observe that the random tropes results are 1 except in the case of the similes median, which is 2. Commonly accepted rhetorical figures mode is 6 and the median is 7. The mode of generated tropes of different categories is always 1 while the median results range between 2 and 3. The mode for generated tropes of same category is between 5 and 7, and the median is between 4 and 5.

Concrete concepts results are shown in Figure 2. Similarly to the abstract concepts, both the results of the mode and the median of random tropes are 1. The mode and the median of commonly accepted rhetorical figures range between 5 and 7. Generated tropes mode of different categories is 1, except for the analogies, which is 7. The median in this case is between 2 and 4. Generated rhetorical figures median of the same category is always 4, while mode is 1 for metaphors, 6 for similes and 7 for analogies.

In Figure 3 the total results for all the rhetorical figures can be seen. Clearly, the result of the randomly generated rhetorical figures is 1. Commonly accepted tropes mode is 7, while the median is 5 for metaphors and similes, and 6 for analogies. The mode of generated tropes of different categories is 1 and 2, and the median is 2 for metaphors, 3 for similes and 4 for analogies. Generated tropes of the same category mode is 7 for metaphors and analogies, while simile mode is 6. The median is 5 for metaphors and similes, and 4 for analogies.

We can conclude that, although the process we have used to generate the rhetorical figures works quite well when concepts of the same category are used, according to the opinions of the evaluators, something different happens in the case of using concepts that belong to different categories, which, in general, obtain worse results. This fact points to the need of using additional properties or relationships in order to obtain concepts that can subsequently give rise to more meaningful rhetorical figures.

Discussion

As we can see, in all cases the randomly generated metaphors are rated as meaningless by the evaluators. In contrast, commonly accepted metaphors get the highest results, with a slight preference for the metaphors created using abstract concepts over the ones that are based on the use of concrete concepts. The automatically generated metaphors using concepts of different categories are also poorly rated, which points out that sharing only one property is not enough to generate a good metaphor. For the generated metaphors using concepts that belong to the same category, the difference that exists between the modes of the metaphors that use concrete and abstract concepts is remarkable. This suggests that abstract metaphors are more evocative and offer a wider range of interpretations than concrete ones. Finally, the overall median for the metaphors also suggests that more aspects need to be taken into consideration to increase the perceived quality of these rhetorical figures.

Table 3 shows the results for the evaluated similes. The ratings in this case are quite similar to the results obtained for the metaphors.

The results of the evaluation of the analogies can be seen in Table 4. The ratings in this case are slightly higher than in the two previous tropes, probably due to the fact that the aspect in which the two concepts are considered to be similar is explicitly stated. This same aspect may be the cause for the lower score obtained by the automatically generated analogies using abstract concepts that belong to the same category. In this case, the similarity perceived by the evaluators may be focused on a different characteristic than the one chosen by the system, which causes the score to be lower than the one granted to the previous figures. On the contrary, the analogies generated using concrete concepts belonging to different categories are much better rated than in the previous tropes. In this case, the reason seems to be the fact that the property used by the system to compare both concepts has been made explicit, so the evaluators can see the reason why the system considers the two concepts related to each other and they are more inclined to accept it as valid.

Finally, the overall results of the evaluation can be seen in Table 5. Although they don't differ much from the results obtained for the different tropes independently, the values of the modes are clearly shifted towards the limits of the scale. This effect suggests that human evaluators tend to accept or not accept a rhetorical figure as valid, but intermediate positions are less common. As for the value of the medians, the condensed results confirm the perception that, in terms of automatically generated tropes, the ones that use abstract concepts that belong to the same category are slightly better appreciated than the rest.

Conclusions and Future Work

We have proved that it is possible to evaluate the quality of rhetorical figures and get consistent results. One of the clearest conclusions is that in our system concepts generate tropes of the same category with significantly higher quality than the tropes based on concepts of different categories.

In view of the results, one of the paths we have to follow

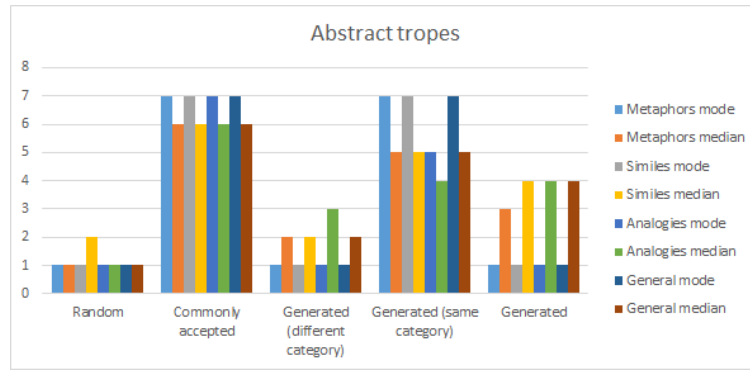


Figure 1: Abstract Tropes

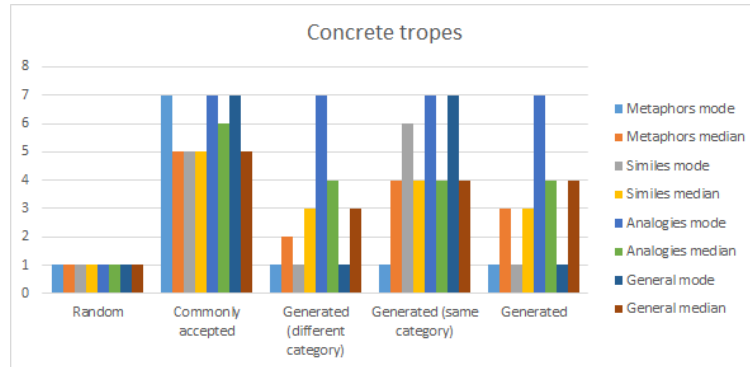


Figure 2: Concrete Tropes

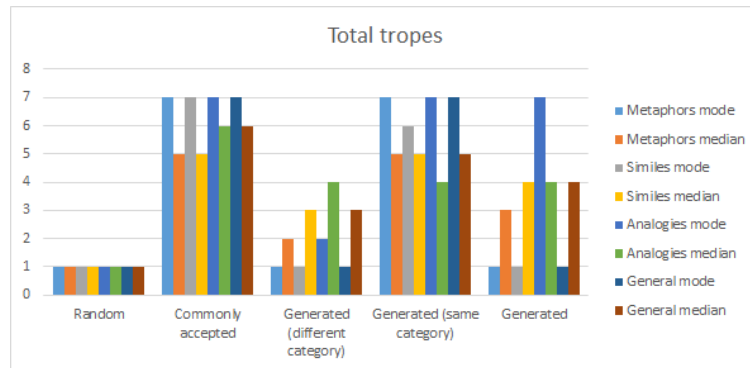


Figure 3: Total Tropes

is directed to find ways to generate good rhetorical figures from concepts of different categories, because in everyday life some of the best rhetorical figures are constructed from these kinds of terms, such as *time is money*. The categories with higher weights obtained for the concept *time* in Thesaurus Rex are *information*, *quantity* and *attribute*, while for the concept *money* they are *thing*, *property*, *value* and *assets*. As we have seen in the evaluation, this trope gets a good rating and we need more information about this type of rhetorical figures.

In order to generate appropriate figurative language depending on the content of a given text, it would be interesting to find sets of words grouped by topic. On the other hand, in order to adequate the figurative language to the goals of the reader, it would be helpful to have sets of concepts grouped by the complexity of their meaning.

Sometimes, constraints encountered arise from the web itself. This is because the information usually available on the web tends to be more literal than figurative. For the previous example, the attributes with more weight obtained from

Thesaurus Rex when searching for *time* are *physical, basic, measurable, relevant and abstract*. That suggests that it may be more appropriate for us to find or generate a specific knowledge resource that provide more evocative properties.

The highest mode for rhetorical figures generated by our system are obtained for analogies. In the case of the median of the total result tendencies are less clear. While the rhetorical figures in the same category produce better results in metaphors and similes, rhetorical figures with different category get better valuations in analogies.

In the future, we would like to continue doing assessments to find patterns or similarities among the best rated rhetorical figures, and we wish to test this with larger datasets. Thus the evaluation findings could serve to improve the quality of the resources generated by our system.

We have used the terms "concrete" and "abstract" when categorising input concepts. It would be interesting to check whether it makes a difference to use a concrete word to describe an abstract concept (e.g. "time is money") and viceversa.

As future work we would also like to check the degree of similarity between the source and target concept. If the concepts are too similar, the resulting trope would be correct but not very practical.

With respect to the amount of information provided in the rhetorical figure, there are no significant differences between those that provide more or less information, because similar results are obtained for metaphors – in which only the original concept and the new concept are indicated – and analogies – in which the shared attribute is also shown.

The results obtained indicate that further attempts should be made to evolve our system and generate higher quality rhetorical figures, progressively evolving the quality of system results towards that of rhetorical figures generated by people. In the future, a useful feature that may improve our system is to relate the original concept with concepts that have more than one property in common. From now on another way that we should investigate is to generate rhetorical figures with concepts that are related through two or more attributes. In the example *A ballerina is a swan*, both concepts share properties as *pretty, graceful and stylized*.

Acknowledgements

This work is partially funded by projects ConCreTe and IDiLyCo. The project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 611733. The IDiLyCo project is funded by the Spanish Ministerio de Economía y Competitividad under grant reference number TIN2015-66655-R.

References

Black, M. 1955. XII. Metaphor. *Proceedings of the Aristotelian Society* 55(1):273–294.

Harmon, S. 2015. Figure8: A novel system for generating and evaluating figurative language. In *Proceedings of the*

Sixth International Conference on Computational Creativity, 71–77.

Hervás, R.; Camara Pereira, F.; Gervás, P.; and Cardoso, A. 2006a. Cross-domain analogy in automated text generation. In *3rd Joint Workshop on Computational Creativity, dentro de la 17th European Conference on Artificial Intelligence*, 43–48.

Hervás, R.; Camara Pereira, F.; Gervás, P.; and Cardoso, A. 2006b. A text generation system that uses simple rhetorical figures. *Procesamiento de Lenguaje Natural* 37:199–206.

Hervás, R.; Costa, R.; Costa, H.; Gervás, P.; and Camara Pereira, F. 2007. Enrichment of automatically generated texts using metaphor. In *6th Mexican International Conference on Artificial Intelligence (MICA-07)*, volume 4827, 944–954. Aguascalientes, Mexico: Springer Verlag, LNAI Series.

Lakoff, G. 1993. The Contemporary Theory of Metaphor. *Metaphor and Thought* 202–251.

Shutova, E.; de Cruys, V.; T.; and Korhonen, A. 2012. Unsupervised Metaphor Paraphrasing using a Vector Space Model. In *Proceedings of the 24th International Conference on Computational Linguistics*, 1121–1130.

Shutova, E.; Sun, L.; and Korhonen, A. 2010. Metaphor identification using verb and noun clustering. In *Proceedings of the 23rd International Conference on Computational Linguistics*, 1002–1010. Beijing: Association for Computational Linguistics.

Veale, T., and Hao, Y. 2011. Exploiting Readymades in Linguistic Creativity: A System Demonstration of the Jigsaw Bard. In *ACL (System Demonstrations)*, 14–19.

Veale, T., and Li, G. 2012. Specifying Viewpoint and Information Need with Affective Metaphors: A System Demonstration of the Metaphor Magnet Web App/Service. In *Proceedings of the ACL 2012 System Demonstrations*, 7–12.

Veale, T., and Li, G. 2013. Creating Similarity: Lateral Thinking for Vertical Similarity Judgments. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 660–670.

Veale, T. 2014a. A Service-Oriented Architecture for Metaphor Processing. In *Proceedings of the Second Workshop on Metaphor in NLP*, 52–60.

Veale, T. 2014b. A service-oriented architecture for metaphor processing. In *Proceedings of the Second Workshop on Metaphor in NLP*, 52–60. Baltimore, MD: Association for Computational Linguistics.

Wallington, A.; Barnden, J.; Buchlovsky, P.; Fellows, L.; and Glasbey, S. 2003. Metaphor annotation: A systematic study. CSRP 03-04, School of Computer Science, The University of Birmingham, Birmingham.

Wilks, Y.; Galescu, L.; Allen, J.; and Dalton, A. 2013. Automatic metaphor detection using large-scale lexical resources and conventional metaphor extraction. In *Proceedings of the First Workshop on Metaphor in NLP*, 36–44. Atlanta, GA: Association for Computational Linguistics.

